Aliasing-free Models of the Glottal Source

Emmanouil Voutetakis

Bachelor's Thesis
Department of Computer Science
University of Crete

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ
UNIVERSITY OF CRETE

Advisor: Dr George P. Kafentzis

December 5, 2024

Contents

List of Figures

List of Tables

Abstract

When a person intends to speak, their brain formulates a sentence with the intended meaning and translates the sequence of words into physiological movements required to produce the corresponding speech sounds. This process involves the contraction of the lungs to push out air through the throat, oral, and nasal cavities. While airflow itself is not audible, sound is created by obstructing airflow to produce oscillations or turbulence. Voiced sounds, a characteristic property of speech signals, are primarily produced when the vocal folds are appropriately tensioned, resulting in oscillations. Other parts of the speech production organs, such as the tongue against the teeth in a rolling "r" or the uvula in the airflow (uvular trill), can also produce oscillations. However, these trills should not be confused with voiced sounds, which are always generated by vocal fold oscillations. Sounds without vocal fold oscillations are classified as unvoiced sounds.

Almost all of the commonly used voice source models are based on piece-wise elementary functions defined in the time domain. The discrete-time implementation of this type of models generally produces aliasing distortion, which is certainly not ideal in most applications and it is something that can be fixed.

This thesis presents the method from Christer Gobl's[1] paper which eliminates the distortion. This method was applied in four models, the Rosenberg C model with some modifications, the LF model, the KLGLOTT88 model with some modifications and the R++ model with some modifications. The complete Laplace transform's of the models are presented, followed by the closed-form expressions of the amplitude spectrum and the phase spectrum for the calculation of the spectrum, with the exception of the KLGLOTT88 and the R++ model where for the open phase of both, Fourier transform was taken immediately.

The central element of the suggested approach revolves around describing the source model in the frequency domain. Through the application of the Laplace transform and phasor arithmetic, it becomes possible to formulate closed-form expressions for the spectrum of the source model. This simplifies the spectrum calculation directly from the model parameters, enabling the derivation of the model's ideal discrete spectrum based on the chosen sampling frequency. This discrete spectrum is completely devoid of aliasing distortion. The sampled glottal flow pulse is then computed using the inverse discrete Fourier transform.

It has to be noted that in all the models except of the LF, the equations have been modified with e^{-at} where a $\rightarrow$ 0, so the signal doesn't change much but with this modification the proposed method can be applied to them.

Chapter 1

Introduction

1.1 The Source-Filter Theory of Speech

Human speech production involves a complex interplay of anatomical components. It begins with the respiratory system, where air from the lungs is expelled. This air passes through the trachea and enters the larynx, where the vocal folds, small muscular folds, are situated. As these folds come together to create a narrow air passage, the airflow causes them to vibrate periodically, generating sound. Above the larynx lies the vocal tract, consisting of various air passages. The shape of the vocal tract, determined by the positions of articulators like the tongue and lips, plays a crucial role in shaping the acoustical characteristics of speech sounds.

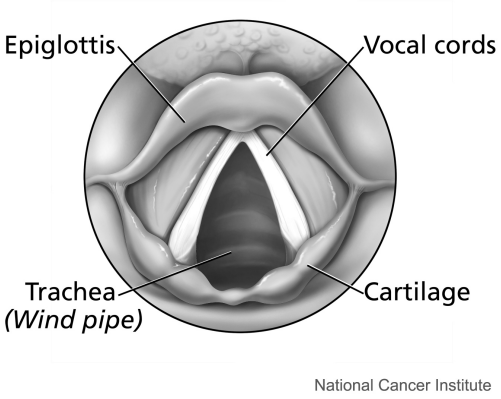


Figure 1.1: Vocal folds as seen from above.[15]

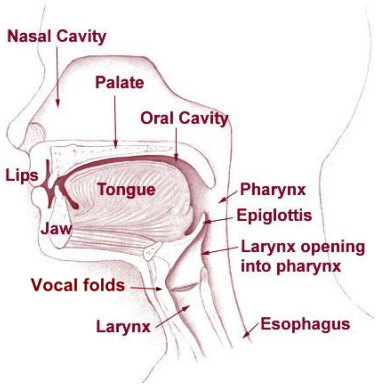


Figure 1.2: Side-view of the speech production organs.[15]

To understand speech production systematically, the source-filter theory divides the process into two stages: (a) The airflow from the lungs causes vibration of the vocal folds, generating the source sound. (b) The spectral characteristics of these source sounds are then shaped by the vocal tract, the filter. The source sound primarily determines vocal pitch, while the filter contributes to the overall spectral structure. The source-filter theory offers a useful approximation of normal human speech, assuming that source sounds are minimally influenced by the vocal tract filter. This theory has found successful application in speech analysis, synthesis, and processing.

The vocal tract can be conceptualized as a tube of varying length, primarily composed of the oral and nasal cavities. While there is a 90-degree bend where the throat transitions into the mouth, the acoustic impact of this bend is typically negligible and can be disregarded in basic models. In terms of acoustic effects, the oral cavity predominantly contributes resonances to the output sound, while the nasal cavities mainly introduce anti-resonances (dips or valleys) in the spectral envelope. Given that the presence of energy is perceptually more significant than its absence, anti-resonances are often overlooked in simplistic models. A straightforward model, therefore, involves a straight cylindrical tube subdivided into constant-radius segments of equal length (see Figure 1.3).

As a much simpler case, consider the acoustic properties of a uniform tube with a length of $L = 17.5$ cm, a standard length for a male vocal tract (see Figures 1.4, 1.5). One end of the tube is closed, representing the glottis, while the other end is open, representing the mouth. Longitudinal sound waves travel inside the tube, moving either towards the mouth or towards the glottis. These waves propagate by compressing and expanding the air in the tube segments, causing the air molecules to displace slightly from their rest positions.

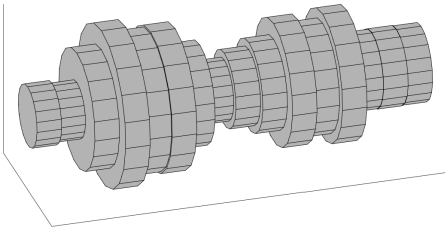


Figure 1.3: Visualization of a vocal tract tube model composed of discrete constant-radius tube sections.[15]

Consequently, the air pressure inside the tube changes over time, depending on the longitudinal displacement of the air along the direction of the traveling wave. Since the air molecules cannot vibrate much at the closed end of the tube, it becomes a point of node. Conversely, the open end of the tube acts as an anti-node, allowing the air molecules to move freely. This arrangement allows for the formation of various standing waves that meet these boundary conditions. Standing waves in a uniform tube with one closed end (the glottis) and one open end (the mouth) exhibit a distinct pattern. Only odd-numbered harmonics are available in such a configuration. 1/4 (purple), 3/4 (green), and 5/4 (sky blue) waves correspond to the first, second, and third resonances ("1/4 wave" means 1/4 of one-cycle waveform is inside the tube).

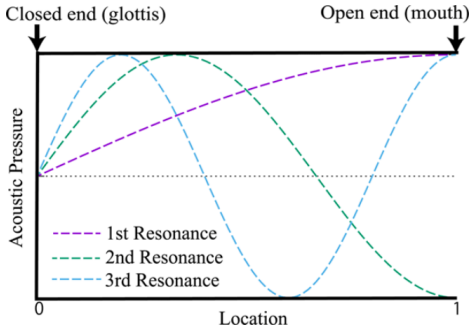


Figure 1.4: Standing waves in a uniform tube[13]

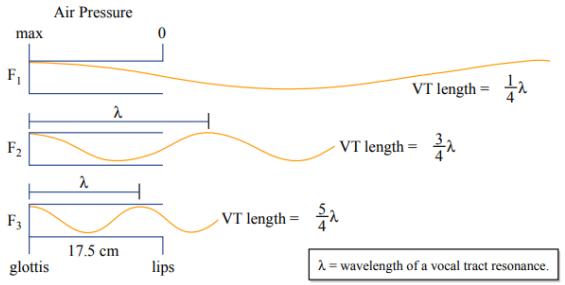


Figure 1.5: Standing waves that meet the boundary conditions[14]

The glottal source originates from the vibrations of the vocal folds. The rate of these periodic glottal vibrations per second is referred to as the fundamental frequency (f_0). In the frequency domain, the glottal source sound dictates the intensities of the fundamental frequency and its harmonics. According to the source-filter theory, the vocal tract functions as an acoustic filter to modify the source sound. Through this filter, specific frequency components are transmitted to the resulting speech, while others are dampened. The properties of this filter are contingent upon the shape of the vocal tract. Comprising the larynx, pharynx, and oral cavities, the vocal tract significantly influences the timbre of the sound. Essentially, its shape determines the resonant frequencies and anti-resonances of the acoustic space, enhancing or diminishing different frequencies of the sound. This shape is influenced by various components, notably the positions of the jaw, lips, and tongue. These resonant frequencies, known as "formants" in phonetics, play a crucial role in shaping the characteristics of speech sounds. When an input frequency component is significantly distant from any of these formants, the vocal tract does not resonate with it. As a result, such frequency components are heavily suppressed, leading to low oscillation amplitudes in the resulting output. Ultimately, the sound wave is emitted from the mouth and nose.

The linear source-filter theory simplifies speech production by treating it as a convolution of the source (vocal fold vibrations and turbulent noise) and the filter (vocal tract resonances). However, this model overlooks the nonlinear nature of speech production, where vocal fold oscillations are influenced by factors like pressure, airflow, tissue elasticity, and vocal tract geometry. In reality, there is a mutual interaction between the source and the filter. The vocal tract influences the source by changing the aerodynamics of glottal flow, and the source sound, propagating through the vocal tract, is partially reflected back to the glottis. This reflection can affect vocal fold oscillations, leading to voice instabilities like sudden pitch jumps, sub-harmonics, resonance, quenching, and chaos. For instance, when the fundamental frequency or its harmonics cross one of the vocal tract resonances, strong acoustic feedback occurs, causing abrupt pitch jumps. This interaction has two aspects: it can facilitate vocal sound production by tuning resonances but can also inhibit vocal fold oscillations, leading

to voice instabilities. Singers must adjust their vocal mechanism to mitigate these instabilities, possibly by modifying the epilarynx to weaken the acoustic coupling during frequency crossings.

The independence between the source and the filter is crucial for effective language-based acoustic communication in humans. This independence allows for the flexible adjustment of the vocal tract configuration to articulate various phonemes sequentially and seamlessly. However, the interaction between the source and the filter can become pronounced when the fundamental frequency or its higher harmonics intersect with one of the vocal tract resonances. This scenario commonly occurs during activities such as singing.

1.2 Introduction to the method

Voice source models are typically formulated in the time domain, described by a collection of piece-wise elementary functions. This convention is observed in various models, including the Liljencrants-Fant model (LF model), the Fujisaki-Ljungqvist model, as well as models presented by Rosenberg (Rosenberg C, R++ models), KLGLOTT88 model, among others.

The voice source signal is inherently a continuous-time signal, and therefore, the functions employed to represent this signal are continuous in time. When translated into a discrete-time system, a common practice is to sample these functions. This involves replacing the continuous variable t with nT_s, where n takes on values of $0, 1, 2, ...$ and T_s denotes the sampling period, which is the reciprocal of the sampling frequency $(1/F_s)$. The act of sampling imposes a constraint on the bandwidth of the sampled waveform, limiting it to the Nyquist frequency, which is half of the sampling frequency. In addition to this inherent limitation, aliasing distortion is introduced due to the source spectrum not being band-limited. Consequently, the discrete-time representation of these continuous-time models will exhibit a distorted frequency spectrum, even within the bandwidth dictated by the sampling frequency.

At extremely high sampling frequencies, this is typically a relatively minor concern. However, in scenarios necessitating a low sampling frequency, it can result in substantial distortion, especially for voices characterized by a high fundamental frequency.

This fold-over distortion, where higher frequencies are perceived as lower frequencies, gives rise to spurious frequency components in the spectrum, except when the sampling frequency coincides with a multiple of the fundamental frequency. In such instances, no additional components are introduced, as each alias component aligns with the frequency of an actual harmonic. Nevertheless, even in this case, the amplitude and phase of the harmonics undergo alteration. While this latter effect is perceptually less disruptive than the introduction of in-harmonic components in the spectrum, it remains sub-optimal for high-precision analysis and modeling of the voice source, especially considering the dynamic nature of fundamental frequency (f_0) in speech and singing which is always changing.

To mitigate this distortion, one strategy is to leverage oversampling of the model waveform. Since the amplitude of the source model spectrum decreases with increasing frequency, oversampling allows for a higher sampling rate. The over-sampled waveform is then subjected to low-pass filtering to diminish the amplitudes of components beyond the Nyquist frequency. Subsequently, the waveform is down-sampled to the desired sampling frequency. However, this approach may not always be the most effective, as achieving optimal results might demand a considerable degree of oversampling. Additionally, while it helps alleviate aliasing, the filtering process may introduce other potential issues. Thus, careful consideration is required when deciding on the appropriate method to address distortion in discrete-time implementations of voice source models.

The conventional approach to discretizing a continuous-time signal involves suppressing frequency components above the Nyquist frequency before signal sampling. This is typically achieved by applying an anti-aliasing filter, namely a low pass filter, in the time domain. Building on these principles, a technique outlined in [10] introduces a method for analytically applying anti-aliasing filtering to pulses described by piece-wise polynomial curves. This filtering is implemented in the time domain before curve sampling. By employing a set of anti-aliasing functions, the source models undergoes modifications in the time domain before sampling, effectively eliminating alias frequency components. However, this process results in a significant attenuation of the original spectral components' amplitudes. Consequently, post-filtering becomes necessary to restore the correct amplitudes. The challenge lies in finding a balance between eliminating aliasing and preserving the integrity of the signal during the discretization process.

This thesis introduces the method from Christer Gobl's[1] paper wherein anti-aliasing filtering is directly conducted in the frequency domain. This frequency domain processing is straightforward and does not necessitate specialized anti-aliasing functions. Moreover, it completely eliminates the need for post filtering since the correct spectrum is preserved up to the Nyquist frequency. However, to make this approach viable, a parametric representation of the model in the frequency domain is essential.

Chapter 2

Aliasing-free Voice Source Modelling

The method suggested for obtaining discrete-time glottal flow waveforms from a continuous-time glottal flow model, without introducing aliasing distortion, is outlined in five main steps. A pre-requisite to understand the methodology requires a short review of complex arithmetic (phasor addition), the Fourier Transform, and the Laplace Transform.

2.1 Phasors

Phasor arithmetic is a powerful mathematical technique used in the analysis of alternating current (AC) circuits, signal processing, and other fields that involve periodic waveforms. It simplifies the manipulation of complex functions by representing them as phasors, which are complex numbers with a specific magnitude and phase angle. Phasor arithmetic involves adding, subtracting, multiplying, and dividing phasors to perform operations on complex functions efficiently.

The addition and subtraction of phasors are straightforward and involve treating them as vectors in the complex plane. For example, let's consider two phasors A and B with magnitudes $|A|$ and $|B|$, and phase angles θ_A and θ_B, respectively. The sum of these phasors $(A+B)$ can be obtained using vector addition in rectangular form:

$$A + B = (|A| \cdot \cos(\theta_A) + |B| \cdot \cos(\theta_B)) + j(|A| \cdot \sin(\theta_A) + |B| \cdot \sin(\theta_B)) \tag{2.1}$$

using Euler's identity:

$$e^{j\theta} = \cos(\theta) + j\sin(\theta) \tag{2.2}$$

Multiplication and division of phasors involve operations on their magnitudes and phase angles. The product of two phasors A and B is given by:

$$A \cdot B = |A| \cdot |B| \cdot e^{j(\theta_A + \theta_B)} \tag{2.3}$$

Similarly, the division of two phasors A and B can be expressed as:

$$\frac{A}{B} = \frac{|A|}{|B|} \cdot e^{j(\theta_A - \theta_B)} \tag{2.4}$$

Phasor arithmetic greatly simplifies the analysis of AC circuits and facilitates the understanding of the relationships between different voltages and currents in a circuit.

Let's consider an example: Suppose we have two phasors $A = 3e^{j(30^\circ)}$ and $B = 2e^{j(45^\circ)}$. The sum of these phasors $A + B$ can be calculated as follows:

$$A + B = 3e^{j(30^\circ)} + 2e^{j(45^\circ)} \tag{2.5}$$

Using the phasor addition formula, we can find the result:

$$A + B = (3 \cdot \cos(30^\circ) + 2 \cdot \cos(45^\circ)) + j(3 \cdot \sin(30^\circ) + 2 \cdot \sin(45^\circ)) \tag{2.6}$$

After evaluating the trigonometric functions, we can express the result in rectangular form, if necessary.

2.2 Laplace Transforms

The Laplace Transform is a powerful mathematical tool used primarily in engineering and physics to analyze linear time-invariant systems. Named after French mathematician Pierre-Simon Laplace, it is employed to transform functions of time, often in the domain of ordinary differential equations, into functions of complex

frequency in the Laplace domain. This transformation simplifies the solving of differential equations, making it a valuable technique for studying dynamic systems. The Laplace Transform of a function $f(t)$ is denoted as $\mathcal{L}\{f(t)\}(s)$ or simply $F(s)$, where s is a complex variable. The Laplace Transform is defined by the integral:

$$F(s) = \mathcal{L}\{f(t)\}(s) = \int_{-\infty}^{\infty} e^{-st} f(t) dt \tag{2.7}$$

One of the key advantages of Laplace Transform is its linearity. This property simplifies the analysis of complex systems composed of multiple elements. Additionally, it turns convolution operations in the time domain into simple multiplication operations in the Laplace domain, further facilitating problem-solving. As an example, consider the step function $u(t)$, defined as:

$$u(t) = \begin{cases} 0 & \text{if } t < 0 \\ 1 & \text{if } t > 0 \end{cases} \tag{2.8}$$

The Laplace Transform of $u(t)$ is given by:

$$\mathcal{L}\{u(t)\}(s) = \frac{1}{s} \tag{2.9}$$

This result is a part of a set of known Laplace Transform pairs that greatly aids in the analysis of various systems.

One critical aspect of the Laplace Transform is the Region of Convergence (ROC), which is the set of values of s for which the integral defining the transform converges. The ROC is essential in determining the applicability and validity of the Laplace Transform for a given function. The Laplace Transform exists only when the complex variable s lies within the ROC.

In applications, the Laplace Transform is frequently used in control theory, signal processing, and circuit analysis. It offers a systematic approach to studying dynamic systems, enabling engineers and scientists to analyze and design systems more effectively.

In Table 2.1 we present some known Laplace Transform pairs. Table 2.2 presents some well-known properties

Function	**Laplace Transform**
$f(t)$	$F(s), \quad R_f$
$u(t)$	$\frac{1}{s}, \quad \Re\{s\} > 0$
$e^{at}u(t)$	$\frac{1}{s-a}, \quad \Re\{s\} > a$
$t^n u(t)$	$\frac{n!}{s^{n+1}}, \quad \Re\{s\} > 0$
$\sin(at)u(t)$	$\frac{a}{s^2+a^2}, \quad \Re\{s\} > 0$
$\cos(at)u(t)$	$\frac{s}{s^2+a^2}, \quad \Re\{s\} > 0$
$u(t)$	$\frac{1}{s}, \quad \Re\{s\} > 0$
$\delta(t-a)$	$e^{-as}, \quad \forall s$
$\sin(at+b)u(t)$	$\frac{a\sin b}{s^2+a^2} + \frac{s\cos b}{s^2+a^2}, \quad \Re\{s\} > 0$
$\cos(at+b)u(t)$	$\frac{a\cos b}{s^2+a^2} - \frac{s\sin b}{s^2+a^2}, \quad \Re\{s\} > 0$
$t^n e^{at}u(t)$	$\frac{n!}{(s-a)^{n+1}}, \quad \Re\{s\} > -a$
$e^{-at}\sin(\omega t)u(t)$	$\frac{\omega}{(s+a)^2+\omega^2}, \quad \Re\{s\} > -a$
$e^{-at}\cos(\omega t)u(t)$	$\frac{s+a}{(s+a)^2+\omega^2}, \quad \Re\{s\} > -a$

Table 2.1: Table of known Laplace Transform pairs.

of the transform.

Property	Laplace Transform
Linearity	$\mathcal{L}\{af(t) + bg(t)\} = a\mathcal{L}\{f(t)\} + b\mathcal{L}\{g(t)\}$
Derivative	$\mathcal{L}\{f'(t)\} = s\mathcal{L}\{f(t)\}$
Integral	$\mathcal{L}\{\int_0^t f(\tau)d\tau\} = \frac{1}{s}\mathcal{L}\{f(t)\}$
Time Shift	$\mathcal{L}\{f(t-a)u(t-a)\} = e^{-as}\mathcal{L}\{f(t)\}$
Frequency Shift	$\mathcal{L}\{e^{at}f(t)\} = \mathcal{L}\{f(t)\}(s-a)$
Scaling in Time	$\mathcal{L}\{f(at)\} = \frac{1}{a}\mathcal{L}\{f(t)\}$
Scaling in Frequency	$\mathcal{L}\{f(t/a)\} = a\mathcal{L}\{f(at)\}$
Convolution	$\mathcal{L}\{f(t) * g(t)\} = \mathcal{L}\{f(t)\} \cdot \mathcal{L}\{g(t)\}$
Time Differentiation	$\mathcal{L}\{t^n f(t)\} = (-1)^n \frac{d^n}{ds^n}\mathcal{L}\{f(t)\}$

Table 2.2: Table of Laplace Transform properties.

2.3 Relationship between Fourier and Laplace Transforms

The Laplace Transform is primarily employed in the analysis of linear time-invariant systems and differential equations, particularly in engineering disciplines. It extends the concept of the Fourier Transform by incorporating complex frequency components, allowing for the analysis of transient and steady-state behavior in dynamic systems. On the other hand, the Fourier Transform focuses solely on decomposing a signal into its frequency components, providing insights into the frequency domain characteristics of a system. The relationship between the two lies in the fact that the Laplace Transform can be seen as an extension or generalization of the Fourier Transform. When the complex frequency parameter in the Laplace Transform is set to be purely imaginary, it reduces to the Fourier Transform, that is, for a signal $x(t)$ with Fourier Transform $X(f)$ and Laplace Transform $X(s)$,

$$X(j\omega) = X(s)\Big|_{s=j\omega} = \int_{-\infty}^{+\infty} x(t)e^{-j\omega t}dt \tag{2.10}$$

2.4 Method for Aliasing-free Glottal Models

The method that will be applied in the rest of the thesis consists of the following steps:

I. In the initial step, the Laplace transform is computed for each segment of the piecewise function that defines the model. Utilizing the linearity property of the Laplace transform, the complete Laplace transform is determined as the sum of the individual transforms.

II. In the second step, for each Laplace transform, the frequency response is derived by evaluating the transform on the imaginary axis by substituting the complex variable s in the Laplace transform with $j\omega$. The resulting frequency response is equivalent to the frequency spectrum of the model. The amplitude response is obtained by calculating the absolute value of the complex frequency response, while the phase response is determined by the angle formed by the real and imaginary parts of the frequency response.

III. In the third step, the method involves deriving closed-form expressions for the overall model spectrum. These expressions play a crucial role in directly calculating the spectrum from the model parameters, an essential aspect of the proposed approach. To achieve this, basic phasor arithmetic is employed, treating each component of the spectrum as a phasor. The process is depicted in Figure 2.1, illustrating the result of adding two phasors, where the length of the phasor represents the amplitude (A), and the angle (relative to the real axis) represents the phase (Φ). The primary formulas for amplitude and phase are presented in equations 2.11 and 2.12, respectively.

IV. In the fourth step, the ideal discrete spectrum of the model is generated from the true model spectrum. This spectrum is essentially a sampled version of the true spectrum up to the Nyquist frequency. The discrete frequency samples are obtained by substituting the continuous frequency variable f in the closed-form expressions of the model with kFs/N for $k = 0, 1, 2..., (N/2) - 1$, where N is an integer greater than the number of samples required to model the glottal pulse. Given that this is a discrete spectrum of a real-valued signal, the complex spectrum should exhibit even symmetric and periodic with a period of F_s.

Consequently, the frequency samples for k between $N/2$ and $N-1$ can be derived from the sample values between 0 and $(N/2)-1$.

V. Once the frequency samples defining the ideal spectrum are available, the corresponding glottal pulse is obtained by computing the Inverse Discrete Fourier Transform (IDFT). It's important to note that the normalization factor of the IDFT should be Fs/N for the correct scaling of the waveform samples.

Moreover, the process described in the last three steps can be extended to multiple glottal pulses by employing phasor arithmetic for individual pulses. This involves treating the spectrum of each glottal pulse as individual phasors. For each subsequent pulse, a linear phase component is added, reflecting the duration of its delay relative to the preceding pulses. For instance, if the duration of the first pulse is T_0, the additional phase component to be added to the phase spectrum of the second pulse is $-2\pi f T_0$.

$$A = \sqrt{\left(A_1 \cos\Phi_1 + A_2 \cos\Phi_2\right)^2 + \left(A_1 \sin\Phi_1 + A_2 \sin\Phi_2\right)^2} = \sqrt{A_1^2 + A_2^2 + 2A_1A_2\cos\left(\Phi_1 - \Phi_2\right)} \tag{2.11}$$

$$\Phi = \operatorname{atan2}\left(A_1 \sin\Phi_1 + A_2 \sin\Phi_2, A_1 \cos\Phi_1 + A_2 \cos\Phi_2\right) \tag{2.12}$$

$$\operatorname{atan2}\left(y, x\right) = \arctan\left(\frac{y}{x}\right) \operatorname{sgn}\left(x\right)^2 + \pi\left(\frac{1 - \operatorname{sgn}\left(x\right)}{2}\right)\left(1 + \operatorname{sgn}\left(y\right) - \operatorname{sgn}\left(y\right)^2\right) \tag{2.13}$$

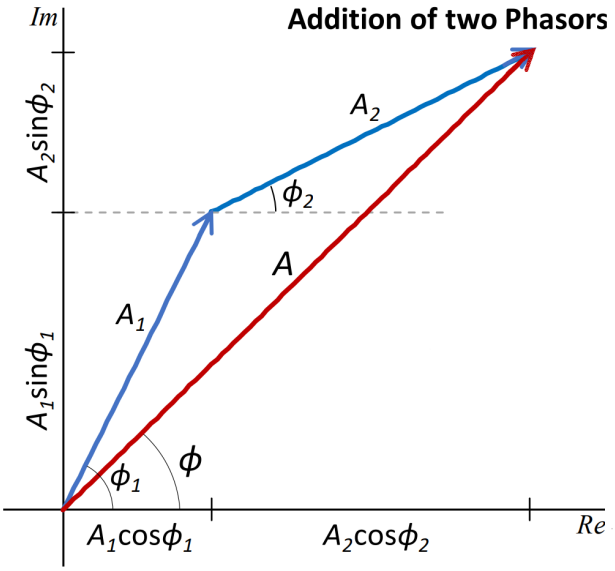


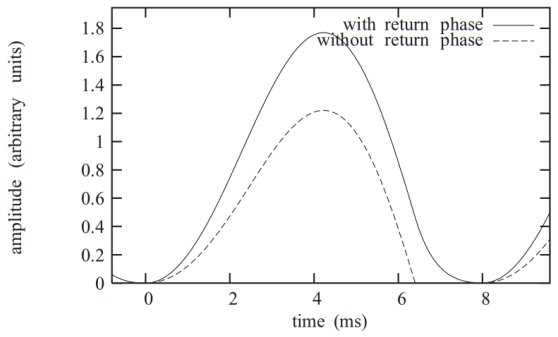
Figure 2.1: The addition of two phasors, where A represents the amplitude response and Φ the phase response of two added components.[1]

Chapter 3

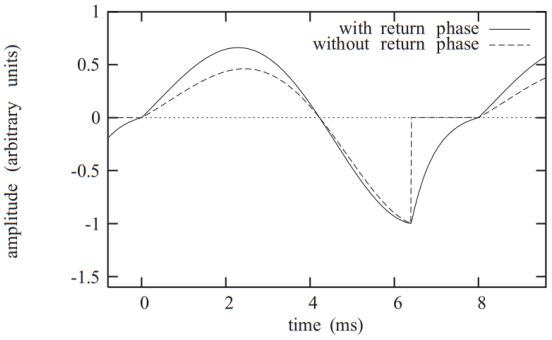
The parameters of the Rosenberg C model

Generating lifelike speech has been a significant area of study for an extended period. Its applications span various fields, including education tools, telecommunications, messaging platforms, accessibility tools, communication aids, language learning platforms, support for hearing-impaired individuals, simulated training environments, technical assistance for intricate tasks, automated announcements, information retrieval systems, and speech translation tools. The pursuit of natural-sounding speech synthesis stands to enhance functionality across these diverse domains.

In physical models, the vibrational behavior of the vocal folds is simulated numerically, relying on a detailed physical representation of the vocal fold system. This approach has been explored by various researchers (Ishizaka, Flanagan, 1972[16]; Liljencrants, 1991[17]; Pelorson, 1994[18]; Herzel, Knudsen, 1995[19]; Story, Titze, 1995[20]; Berry, Titze, 1996[21]; de Vries, 1999[22]; Gunter, 2003[23]). Conversely, parametric models describe the glottal flow or its first derivative using piece-wise analytical functions representing different phases of the glottal cycle. Parametric source models are particularly advantageous for analyzing the source signal and characterizing it efficiently with a concise set of parameters.



(a) GFMs with abrupt (Rosenberg C) or smooth closure (LF).[2]



(b) The derivative of the GFMs with abrupt (Rosenberg C) or smooth closure (LF).[2]

The fundamental concept underlying the Rosenberg C glottal model is to represent the glottal airflow as a series of pulses, with each pulse corresponding to a cycle of opening and closing of the vocal folds. These pulses are commonly characterized using mathematical functions like rectangular or triangular shapes, which approximate the changes in airflow velocity throughout the glottal cycle.

The Rosenberg C model, defined by equation (3.1), comprises a trigonometric function representing the opening and closing phases of the vocal folds. The function consists of two segments: the first segment characterizes the opening phase, while the second represents the closing phase. In the equation, A denotes the amplitude of voicing, which is a constant associated with the amplitude of the glottal pulse. T_0 represents the glottal time interval, T_p denotes the opening time, and $T_n = T_0 - T_p$ represents the closing time of the vocal folds. $T_p = \alpha_1 T_0$ is the opening time and $T_n = T_0 - T_p = \alpha_2 T_0$ is the closing time of the vocal folds, with α_1 and α_2 constants related to the increasing and the decreasing slope of the glottal pulse, respectively[5].
Key features of the Rosenberg C glottal model include:

- **Pulse Shape** : The shape of each pulse represents the variation in airflow velocity during a single glottal cycle. Different pulse shapes can be used depending on the desired level of detail and accuracy in modeling the glottal airflow.

- **Parameters**: The model typically includes parameters to control the duration, amplitude, and shape of the pulses, allowing for flexibility in adjusting the waveform to match empirical data or physiological observations.
- **Periodicity**: By repeating the pulse pattern at regular intervals, the model captures the periodic nature of vocal fold vibration during speech production. The fundamental frequency (pitch) of the speech signal is determined by the rate at which these pulses occur.
- **Physiological Interpretation**: While the Rosenberg C glottal model provides a mathematical representation of glottal airflow, it may not fully capture the complex bio-mechanical processes involved in vocal fold vibration. Nevertheless, it serves as a valuable tool for analyzing and synthesizing speech signals in various applications, such as speech synthesis, voice pathology assessment, and acoustic phonetics research.

The Rosenberg C model is a model of differentiated glottal flow, and the piece-wise function defining the pulse is made up of three parts according to (3.1).

$$U_g'(t) = \begin{cases} \dfrac{\pi A}{2T_p} \sin\left(\dfrac{\pi t}{T_p}\right), & 0 \le t \le T_p \\ -\dfrac{\pi A}{2T_n} \sin\left(\dfrac{\pi}{2}\dfrac{t-T_p}{T_n}\right), & T_p \le t \le T_p + T_n \\ 0, & T_p + T_n \le t \le T_0 \end{cases} \tag{3.1}$$

The Rosenberg C model had to undergo modification for the application of the proposed method. This adjustment was necessary as it was feasible to compute the Laplace transform, but replacing s with $j\omega$ presented a challenge due to the constraint that $\mathrm{Re}\{\mathrm{s}\} > 0$.

The modified version is (3.2) with a $\to 0$:

$$U_g'(t) = \begin{cases} \dfrac{\pi A}{2T_p} e^{-at} \sin\left(\dfrac{\pi t}{T_p}\right), & 0 \le t \le T_p \\ -\dfrac{\pi A}{2T_n} e^{-a(t-T_p)} \sin\left(\dfrac{\pi}{2}\dfrac{t-T_p}{T_n}\right), & T_p \le t \le T_p + T_n \\ 0, & T_p + T_n \le t \le T_0 \end{cases} \tag{3.2}$$

The first two segments models the flow derivative during the open phase of the glottal cycle. The open phase lasts for a duration of T_0 and concludes when the waveform's amplitude reaches: $-\frac{\pi A}{2T_n}$, representing the negative amplitude of the primary glottal excitation. There is no return phase because Rosenberg C has abrupt closure. Here are the fundamental parameters for the Rosenberg C model outlined below.

Variable	Description
A	amplitude
T_0	fundamental period
T_p	maximum of the glottal flow waveform
T_n	time interval between maximum of the glottal flow waveform and the GCI

Table 3.1: Rosenberg C variables and Descriptions

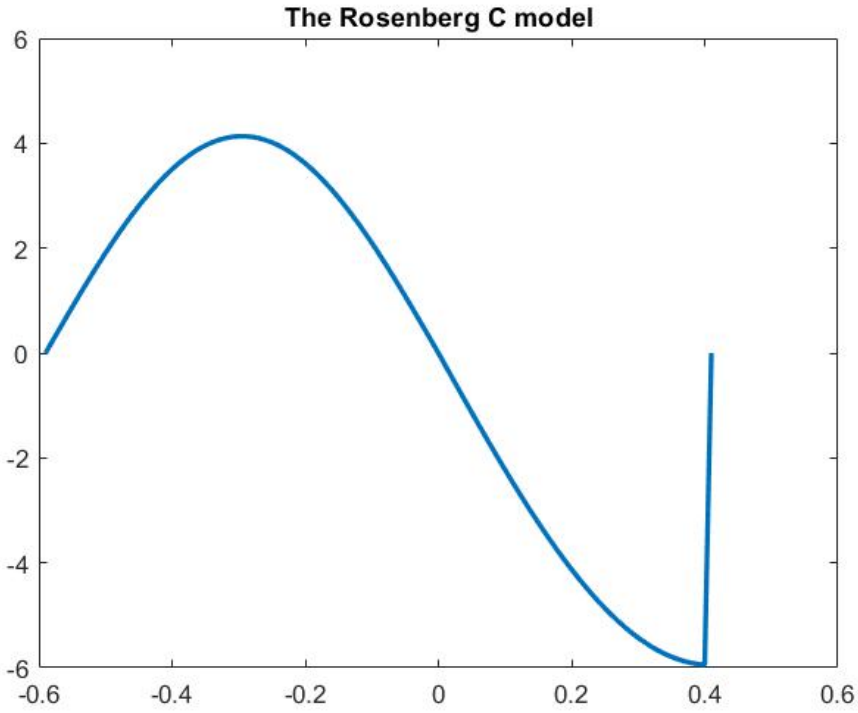


Figure 3.2: The Rosenberg C pulse

Chapter 4

The parameters of the LF model

The LF-model is well-suited for non-interactive flow parameterization as it offers a comprehensive fit to typical wave shapes using a minimal parameter set. It is also adaptable to accommodate extreme phonations while maintaining simplicity in digital implementations, despite potential complexities arising from parameter interdependencies.

The LF-model offers a parameterized representation of the fundamental characteristics of the glottal flow derivative's coarse structure. This structure predominantly reflects the dynamics and dimensions of the glottis, alongside pulse skew attributed to the influence of the vocal tract. Key aspects targeted by this coarse structure include the open quotient, the rate of opening and closing, and the correlation between the glottal pulse and the peak glottal flow. The open quotient, which varies among speakers, empirically adjusts the relative amplitudes of initial harmonics. Typically, breathy voices exhibit larger open quotients, whereas pressed voices show smaller values. The efficiency of a speaker is reflected in the relationship between the peak glottal flow and the amplitude

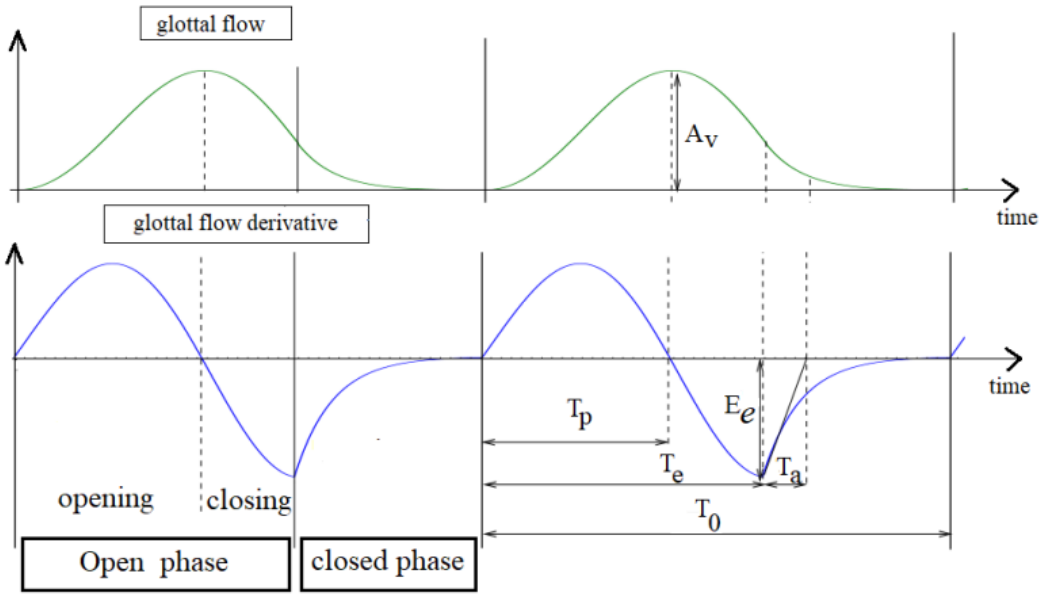


Figure 4.1: Example of the LF glottal pulse and its derivative [5].

of the glottal pulse. While the glottal pulse serves as the primary excitation for voiced speech, it's actually the slope of the glottal flow at closure, rather than the peak glottal flow, that predominantly influences the speaker's loudness. Ripple, if timed appropriately with supra-glottal pressure peaks coinciding with the glottal flow, can also affect efficiency by attenuating the glottal flow without impacting the closure rate. The model utilized is described by the following equation (4.1), with parameters E_0 (4.2) and T_a (4.3):

$$U'_g(t) = \begin{cases} E_0 e^{at} \sin\left(\dfrac{\pi t}{T_p}\right), & 0 \le t \le T_e \\ -\dfrac{E_e}{\varepsilon T_a}\left(e^{-\varepsilon(t-T_e)} - e^{-\varepsilon(T_0-T_e)}\right), & T_e \le t \le T_0 \end{cases} \tag{4.1}$$

where

$$E_0 = -E_e \frac{e^{-aT_e}}{\sin\left(\frac{\pi T_e}{T_p}\right)} \tag{4.2}$$

and

$$T_a = \frac{\left(1 - e^{-\varepsilon(T_0-T_e)}\right)}{\varepsilon} \tag{4.3}$$

The model is characterized by four parameters. Three parameters, E_0, T_p, and a, delineate the open phase, while T_a represents the return phase. Maintaining continuity between the open and return phases at the point T_e, ε is reliant on T_a. Although the relationship between ε and T_a lacks a closed-form expression, for small values of T_e, ε approximates T_a. Among these parameters, T_a holds significant importance in terms of human perception, as it governs the extent of spectral tilt evident in the source.

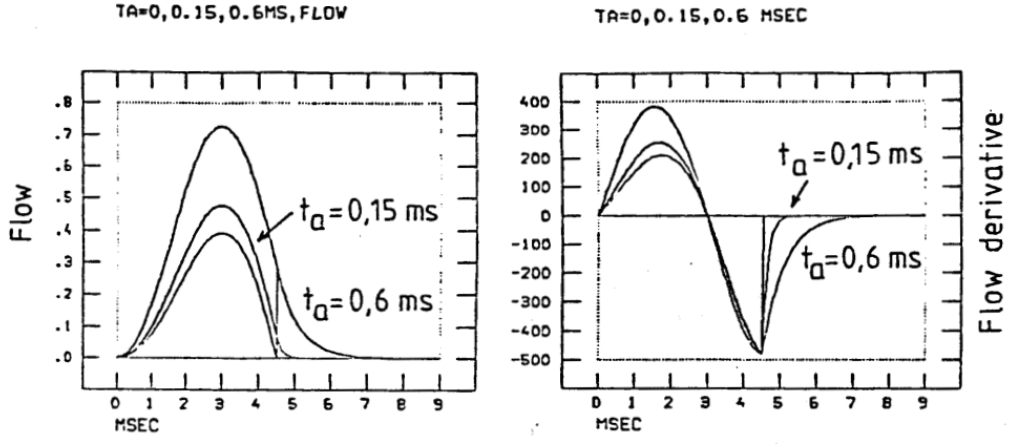


Figure 4.2: Example of the LF glottal pulse and its derivative with different T_a values [3].

The LF-model represents the glottal flow derivative. Given that the glottal flow must be zero at both $t = 0$ and $t = T_0$, the integral of the glottal flow derivative from 0 to T_0 must also be zero. This necessary condition elucidates the implicit equation for a. Apart from the constant factor $-E_e$, the left side of the equation denotes the integral of $U'_g(t)$ from 0 to T_e, while the right side represents the negative integral of $U'_g(t)$ from T_e to T_0.

The implicit equation for ε stems from the wave-form's continuity at T_e. Additionally, it's feasible to derive an expression for ε as a function of a and the model parameters by combining the two equations. The open quotient is defined as:

$$O_q = \frac{T_e}{T_0} \tag{4.4}$$

Given that this definition doesn't consider the duration of the return phase, Fant proposed the following modification:

$$O_q = \frac{T_e + T_a}{T_0} \tag{4.5}$$

The LF model offers several advantages:

- **Parameterization of Glottal Flow**: It provides a parameterized representation of the glottal flow derivative, allowing for the characterization of key aspects of vocal fold vibration and airflow during speech production.
- **Capture of Essential Features**: The model captures essential features such as the open quotient, speed of opening and closing, and the relationship between the glottal pulse and the peak glottal flow, which are important for synthesizing natural-sounding speech.
- **Computational Efficiency**: The LF model strikes a balance between computational efficiency and accuracy, making it suitable for real-time speech synthesis applications and simulations where efficiency is critical.

- **Widespread Use**: It has been widely adopted in both academic research and commercial speech synthesis systems due to its effectiveness and simplicity.
- **Versatility**: The LF model can be adapted and extended to accommodate different linguistic contexts, speaker characteristics, and speech styles, making it versatile for various speech synthesis tasks.

The LF model is a model of differentiated glottal flow, with the pulse defined by a piece-wise function comprising two parts, as outlined in equation (4.6)

$$U'_g(t) = \begin{cases} -E_e e^{a(t-T_e)} \dfrac{\sin\left(\frac{\pi t}{T_p}\right)}{\sin\left(\frac{\pi T_e}{T_p}\right)}, & 0 \le t \le T_e \\ -\dfrac{E_e}{\varepsilon T_a}\left(e^{-\varepsilon(t-T_e)} - e^{-\varepsilon(T_0-T_e)}\right), & T_e \le t \le T_0 \end{cases} \tag{4.6}$$

In the initial segment, the flow derivative is modeled during the open phase of the glottal cycle using a sine function that experiences exponential growth in amplitude. The open phase spans a duration of T_e and concludes at time-point T_e when the wave-form's amplitude reaches $-E_e$, representing the negative amplitude of the primary glottal excitation. The subsequent segment describes the flow derivative after the main excitation, encompassing the return phase of the pulse.

Here are the fundamental parameters for the LF model outlined below.

Variable	Description
E_e	amplitude at the minimum of the glottal flow derivative (maximum of excitation)
T_0	fundamental period
T_e	instant of maximum excitation
T_p	instant of the maximum of the glottal flow
T_a	time constant of the return phase

Table 4.1: LF variables and Descriptions

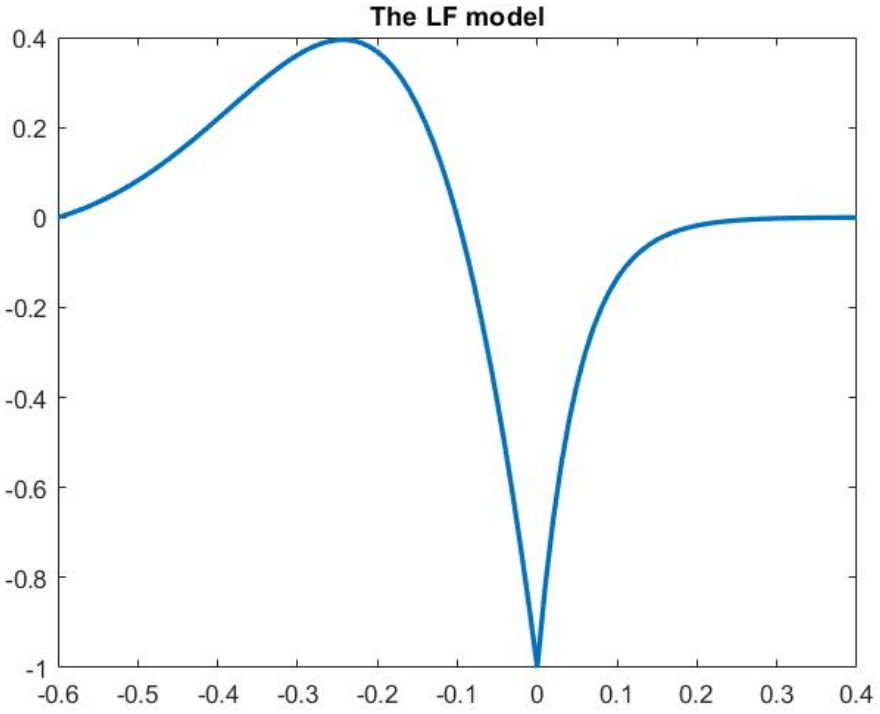


Figure 4.3: The LF pulse.

Overall, the LF model provides a practical and effective framework for modeling and synthesizing speech, contributing to advancements in speech technology and human-computer interaction.

Chapter 5

The parameters of the KLGLOTT88 model

This model introduces several source control parameters:

Variable	Description
A_V	Amplitude of voicing, measured in dB
F_0	Voicing fundamental frequency, expressed in Hz
O_q	Open quotient of the glottal waveform, represented as a percentage of a full period
TL	Tilt of the voicing source spectrum, indicated as dB down at 3 kHz
FL	Period-to-period flutter (quasirandom fluctuations) information, measured as a percentage of maximum
DI	Degree of diplophonic double-pulsing irregularity information, also as a percentage of maximum
AH	Amplitude of aspiration (breathiness) noise, in dB

Table 5.1: KLGLOTT88 variables and Descriptions

We are going to be using only the fundamental parameters, which are T_0, A_V and O_q. During the open phase of a glottal cycle, the volume velocity waveform follows a parametric relationship initially proposed by Rosenberg (1971)[8]. This relationship is expressed as:

$$U_g(t) = at^2 - bt^3 \tag{5.1}$$

where a (5.2) and b (5.3) are constants determined by the amplitude of voicing and the duration of the open period.

$$a = \left(\frac{27}{4}\right)\left(\frac{A_V}{(O_q^2 T_0)}\right) \tag{5.2}$$

$$b = \left(\frac{27}{4}\right)\left(\frac{A_V}{(O_q^3 T_0^2)}\right) \tag{5.3}$$

In time-domain models, glottal flow models are generally not viewed as filters or linear systems; instead, they are Characterized by empirical or arbitrary equations, often based on polynomials or trigonometric functions. An example of such a model is the KLGLOTT88 model proposed by D. Klatt and L. Klatt (1990)[6]. This model is described by the following equation (5.4) when there is no additional spectral tilt component:

$$U_g(t) = \begin{cases} at^2 - bt^3, & 0 \le t \le O_q T_0 \\ \\ 0, & O_q T_0 \le t \le T_0 \end{cases} \tag{5.4}$$

The corresponding impulse response, magnitude and phase spectra of are plotted in Figure 5.1.

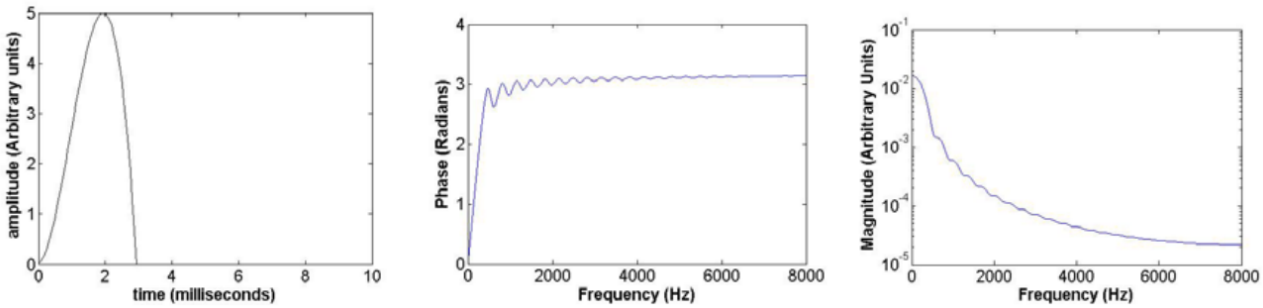


Figure 5.1: KLGLOTT88 glottal flow model, as assumed by the Klatt synthesizer (D. Klatt and L. Klatt, 1990). From left to right: glottal waveform, spectral phase and spectral magnitude.[7]

The spectral impact of different open quotients, spectral tilts, and aspiration levels is illustrated in Figure 5.2. The aspiration component introduces noise energy with a relatively uniform spectrum, as depicted in the final panel of the figure.

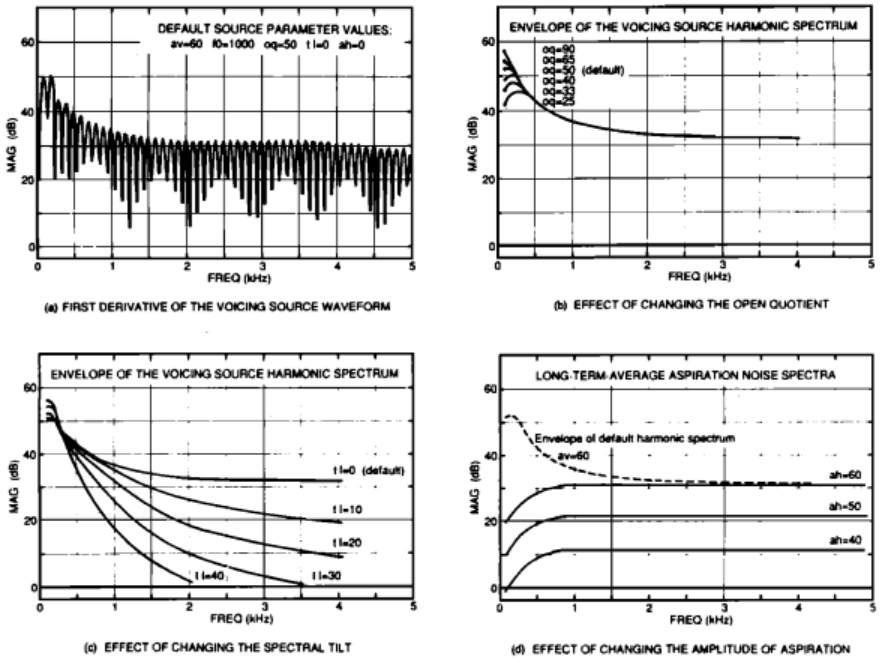


Figure 5.2: The magnitude spectra of $U_g'(t)$ (where $U_g(t)$ is modified by a first difference approximation to the radiation characteristic) synthesized by the KLGLOTT88 voicing source model are depicted at various values for each of three control parameters. Panel (a) illustrates the spectrum of a pulse train with a fundamental frequency of $f_0 = 100Hz$, while panels (b) and (c) demonstrate the spectral envelope of a harmonic spectrum generated by synthesizing such a pulse train. Panel (d) displays the spectra of the aspiration noise source [6].

Present acoustic models of the $U_g(t)$ waveform are relatively simplistic, capturing only the fundamental shapes and spectral attributes evident in natural glottal waveforms. While these models show potential for speech synthesis applications, further modeling intricacies might be required to attain high quality synthesis across both male and female voices.

The KLGLOTT88 model, derived from the Rosenberg B model, has been utilized in various applications such as the Klatt synthesizer and several research studies [6]. The glottal waveform, denoted as U_g, is defined by four essential parameters: the fundamental frequency ($F_0 = 1/T_0$), the amplitude of voicing (A_V), the open quotient (O_q), and the attenuation of a spectral tilt filter (TL). When the spectral tilt filter attenuation parameter TL is set to 0 dB, the KLGLOTT88 model coincides with the Rosenberg B model. In this scenario,

the equations governing the model and its derivative are as follows:

$$U'_g(t) = \begin{cases} 2at - 3bt^2, & 0 \le t \le O_q T_0 \\ \\ 0, & O_q T_0 \le t \le T_0 \end{cases} \tag{5.5}$$

The KLGLOTT88 model had to undergo modification for the application of the proposed method. This adjustment was necessary as it was feasible to compute the Laplace transform, but replacing s with $j\omega$ presented a challenge due to the constraint that Re{s} > 0.

For convenience, the variable a has been changed to g because the variable a was required for the modification described below:
The modified version is (5.6) with a $\to 0$:

$$Ug'(t) = \begin{cases} 2ge^{-at}t - 3be^{-at}t^2, & 0 \le t \le O_q T_0 \\ \\ 0, & O_q T_0 \le t \le T_0 \end{cases} \tag{5.6}$$

Also we have:

$$g = \left(\frac{27}{4}\right)\left(\frac{A_V}{(O_q^2 T_0)}\right) \tag{5.7}$$

$$b = \left(\frac{27}{4}\right)\left(\frac{A_V}{(O_q^3 T_0^2)}\right) \tag{5.8}$$

Here are the fundamental parameters for the KLGLOTT88 model outlined below.

Variable	**Description**
T_0	fundamental period
AV	amplitude of voicing
O_q	the open quotient

Table 5.2: KLGLOTT88 variables and Descriptions

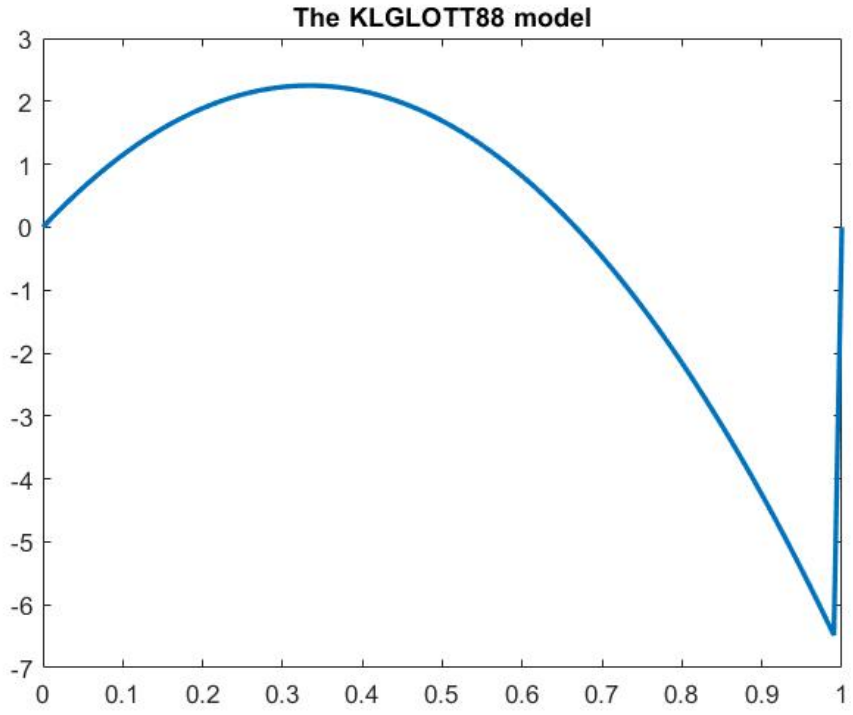


Figure 5.3: The KLGLOT88 pulse

Chapter 6

The parameters of the R++ model

R. Veldhuis presents an alternative model for the LF-model, derived from the Rosenberg B model, called the R++ model (Rosenberg, 1971[8], Fant et al., 1985[3]). The R++ model shares similar features with the LF-model but offers computational efficiency advantages. This enhanced model incorporates improvements along two axes: an asymmetry coefficient and a return phase parameter. The pulse waveform, defined by a piece-wise function, comprises three distinct parts according to (6.1).

The computational efficiency of the R++ model stands out as a significant advantage, allowing for faster processing while maintaining high-quality synthesized speech. This efficiency is especially advantageous in real-time applications where rapid performance is crucial, including speech recognition systems, voice assistants, and telecommunications platforms.

$$U'_g(t) = \begin{cases} 4Kt(T_p - t)(T_x - t), & 0 \le t \le T_e \\ U'_g(T_e)\left(\dfrac{e^{-\frac{t-T_e}{T_a}} - e^{-\frac{T_0-T_e}{T_a}}}{1 - e^{-\frac{T_0-T_e}{T_a}}}\right), & T_e \le t \le T_0 \end{cases} \tag{6.1}$$

The R++ model had to undergo modifications for the application of the proposed method. This adjustment was necessary as it was feasible to compute the Laplace transform, but replacing s with $j\omega$ presented a challenge due to the constraint that Re{s} > 0.

The modified version is (6.2) with a $\to 0$:

$$U'_g(t) = \begin{cases} 4Ke^{-a(t+T_e)}(t + T_e)(T_p - t - T_e)(T_x - t - T_e), & -T_e \le t \le 0 \\ U'_g(0)\left(\dfrac{e^{-\frac{t}{T_a}} - e^{-\frac{T_0-T_e}{T_a}}}{1 - e^{-\frac{T_0-T_e}{T_a}}}\right), & 0 \le t \le T_0 - T_e \end{cases} \tag{6.2}$$

Solving for T_x, the following expression is obtained:

$$T_x = T_e\left(1 - \frac{0.5T_e^2 - T_eT_p}{2T_e^2 - 3T_eT_p + 6T_a(T_e - T_p)D(T_0, T_e, T_a)}\right) \tag{6.3}$$

The factor $D(T_0, T_e, T_a)$ appears as a consequence of the exponential return phase, as defined in the LF-model.

$$D(T_0, T_e, T_a) = 1 - \frac{(T_0 - T_e)/T_a}{\exp\left(\frac{T_0-T_e}{T_a}\right) - 1} \tag{6.4}$$

Here are the fundamental parameters for the R++ model outlined below.

Variable	**Description**
K	amplitude coefficient
T_0	fundamental period
T_e	minimum of the glottal flow derivative waveform (excitation instant)
T_p	maximum of the glottal flow waveform
T_a	time constant for the return phase

Table 6.1: R++ variables and Descriptions

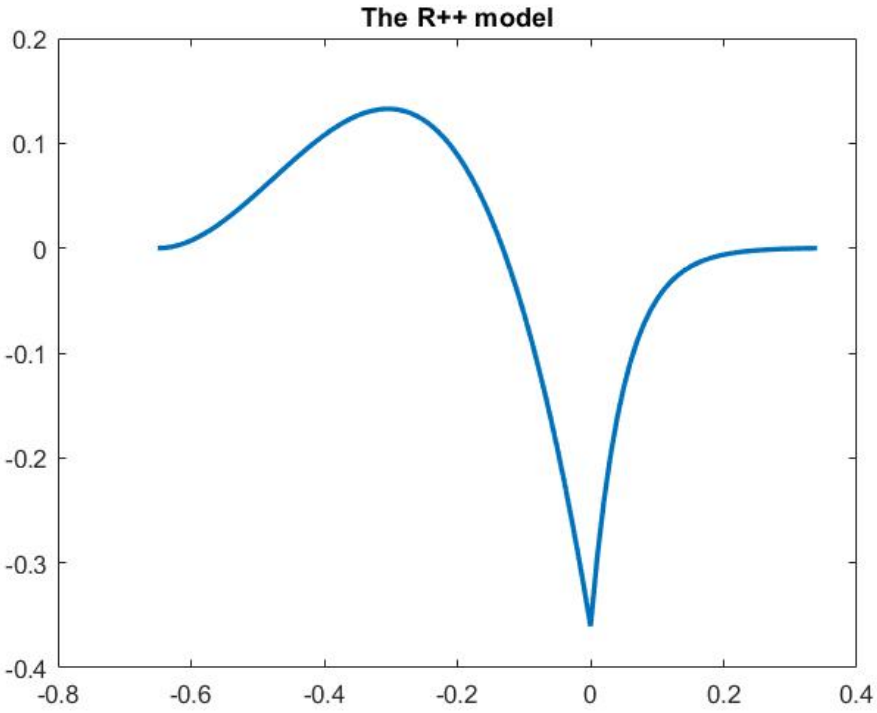


Figure 6.1: The R++ pulse

Overall, the R++ model represents a significant advancement in speech synthesis technology, offering improved performance and versatility compared to its predecessors. Its development reflects ongoing efforts to enhance the accuracy, naturalness, and efficiency of synthetic speech generation methods.

Chapter 7

The Rosenberg C model in the frequency domain

The Rosenberg C model is a mathematical representation of the glottal flow waveform, specifically designed to capture the characteristics of vocal fold vibration during speech production. In this model, the glottal flow is expressed as a piece-wise function, consisting of two parts that represent the glottal cycle. The Rosenberg C model provides a mathematical framework to describe the glottal flow waveform in the time domain, capturing essential features of vocal fold dynamics during speech production.

The first part of the model corresponds to the anti-causal response, while the second part represents the causal response. Doing that modification the $U_g'(t)$ is now:

$$U_g'(t) = \begin{cases} \dfrac{\pi A}{2T_p} e^{-a(t+T_p)} \sin\left(\dfrac{\pi}{T_p}(t+T_p)\right), & -T_p \le t \le 0 \\ -\dfrac{\pi A}{2T_n} e^{-at} \sin\left(\dfrac{\pi}{2}\dfrac{t}{T_n}\right), & 0 \le t \le T_n \\ 0, & T_n \le t \le T_0 \end{cases} \tag{7.1}$$

The Laplace transform of the first part of the Rosenberg C model is shown in (7.2), where $u(t)$ is the unit step function.

$$H_1(s) = \mathcal{L}\left\{\left[\frac{\pi A}{2T_p} e^{-a(t+T_p)} \sin\left(\frac{\pi}{T_p}(t+T_p)\right)\right][u(t+T_p) - u(t)]\right\} = \frac{\pi^2 A}{2T_p^2} \frac{e^{sT_p} + e^{-aT_p}}{(s+a)^2 + (\frac{\pi}{T_p})^2} \tag{7.2}$$

The Laplace transform of the second part of the Rosenberg C model is shown in (7.3), where $u(t)$ is the unit step function.

$$H_2(s) = \mathcal{L}\left\{\left[-\frac{\pi A}{2T_n} e^{-at} \sin\left(\frac{\pi}{2}\frac{t}{T_n}\right)\right][u(t) - u(t-T_n)]\right\} = -\frac{\pi A}{2T_n} \frac{\frac{\pi}{2T_n} - (s+a)\, e^{-T_n(s+a)}}{(s+a)^2 + (\frac{\pi}{2T_n})^2} \tag{7.3}$$

Since the Rosenberg C pulse is the sum of the two parts of the model, the complete transform for the Rosenberg C model is:

$$H(s) = H_1(s) + H_2(s) \tag{7.4}$$

Next, we substitute s with $j\omega$ in equation (7.4) to obtain the corresponding frequency response, which is essentially the frequency spectrum of the model. The absolute value of $H(j\omega)$ gives the amplitude spectrum: $A_1(f)$ and $A_2(f)$ for the first part (7.5) and the second part (7.7), respectively. The phase spectrum is determined from the angle formed by the real and imaginary parts of the transform: $\Phi_1(f)$ and $\Phi_2(f)$ for the first part (7.6) and the second part (7.8), respectively. Also $Re_1(f)$, $Im_1(f)$, $Re_2(f)$, $Im_2(f)$ are (7.9), (7.10), (7.11), (7.12) respectively.

$$A_1(f) = \frac{\sqrt{Re_1(f)^2 + Im_1(f)^2}}{\sqrt{\left(\left(\frac{\pi}{T_p}\right)^2 + a^2 - (2\pi f)^2\right)^2 + (4a\pi f)^2}} \tag{7.5}$$

$$\Phi_1(f) = \operatorname{atan2}(Im_1(f), Re_1(f)) - \operatorname{atan2}\left(4a\pi f, a^2 + \left(\frac{\pi}{T_p}\right)^2 - (2\pi f)^2\right) \tag{7.6}$$

$$A_2(f) = \frac{\sqrt{Re_2(f)^2 + Im_2(f)^2}}{\sqrt{\left(\left(\frac{\pi}{2T_n}\right)^2 + a^2 - (2\pi f)^2\right)^2 + (4a\pi f)^2}} \tag{7.7}$$

$$\Phi_2(f) = \text{atan2}(Im_2(f), Re_2(f)) - \text{atan2}\left(4a\pi f, a^2 + \left(\frac{\pi}{2T_n}\right)^2 - (2\pi f)^2\right) \tag{7.8}$$

$$Re_1(f) = \frac{\pi^2 A}{2T_p^2}\left(e^{-aT_p} + \cos(2\pi f T_p)\right) \tag{7.9}$$

$$Im_1(f) = \frac{\pi^2 A}{2T_p^2}\sin(2\pi f T_p) \tag{7.10}$$

$$Re_2(f) = -\frac{\pi A}{2T_n}\left(\frac{\pi}{2T_n} - e^{-aT_n}(2\pi f \sin(2\pi f T_n) + a\cos(2\pi f T_n))\right) \tag{7.11}$$

$$Im_2(f) = -\frac{\pi A}{2T_n}e^{-aT_n}(a\sin(2\pi f T_n) - 2\pi f \cos(2\pi f T_n)) \tag{7.12}$$

By employing the phasor arithmetic outlined in Chapter 2, we obtain the comprehensive Rosenberg C model spectrum as illustrated in equations (7.13) and (7.14), with $atan2$ defined according to 2.13:

$$A = \sqrt{A_1^2 + A_2^2 + 2A_1A_2\cos(\Phi_1 - \Phi_2)} \tag{7.13}$$

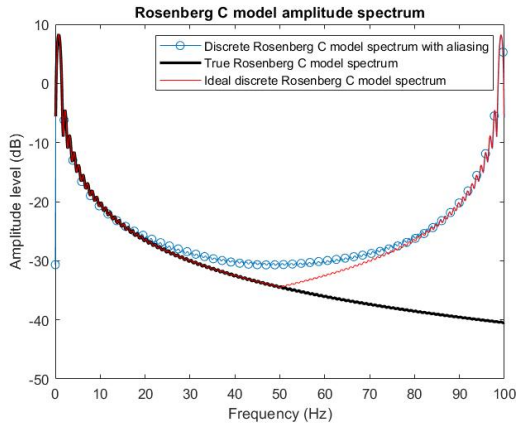
$$\Phi = \text{atan2}\left(A_1\sin\Phi_1 + A_2\sin\Phi_2, A_1\cos\Phi_1 + A_2\cos\Phi_2\right) \tag{7.14}$$

To shift the pulse to start at time-point $t_0 = 0$ rather than at T_p, a phase delay of $2\pi f T_p$ should be added to (7.14), so now Φ is (7.15):

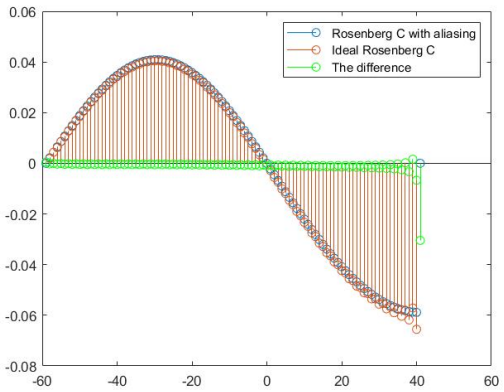
$$\Phi = \text{atan2}\left(A_1\sin\Phi_1 + A_2\sin\Phi_2, A_1\cos\Phi_1 + A_2\cos\Phi_2\right) - 2\pi f T_p \tag{7.15}$$

After evaluating them and composing the signal and applying IFFT on matlab we get the following results.

(As mentioned previously, the Rosenberg C model was modified to make the method applicable to it.)



(a) Amplitude spectrum of the Rosenberg C model showing the true spectrum (black), ideal discrete spectrum (red) and aliased discrete spectrum (blue). $F_s = 100Hz$



(b) Rosenberg C pulse using the proposed method (red), compared to the corresponding sampled Rosenberg C pulse (blue). The difference between the two is shown by the green line.

Chapter 8

The LF model in the frequency domain

When developing the frequency domain representation of the LF-model, it is advantageous to consider the LF pulse as the non-causal response to an impulse with an amplitude of -E_e at the time-point T_e. Consequently, we set $t = 0$ at the time of the main excitation (T_e), deviating from the conventional practice of setting it at the time of glottal opening (T_0). In this representation, the first part of the model (the open phase) corresponds to the anti-causal response, while the second part (the return phase) signifies the causal response. Doing that modification the $U'_g(t)$ is now:

$$U'_g(t) = \begin{cases} -E_e e^{at} \dfrac{\sin\left(\frac{\pi}{T_p}(t+T_e)\right)}{\sin\left(\frac{\pi T_e}{T_p}\right)}, & -T_e \le t \le 0 \\ -\dfrac{E_e}{\varepsilon T_a}\left(e^{-\varepsilon t} - e^{-\varepsilon(T_0-T_e)}\right), & 0 \le t \le T_0 - T_e \end{cases} \tag{8.1}$$

The Laplace transform of the first part of LF model (the open phase) is shown in (8.2), where u(t) is the unit step function.

$$H_1(s) = \mathcal{L}\left\{\left[-\frac{E_e}{\sin\left(\frac{\pi T_e}{T_p}\right)}\sin\left(\frac{\pi}{T_p}(t+T_e)\right)\right][u(t+T_e) - u(t)]\right\} \tag{8.2}$$

$$= E_e \frac{s - a + \frac{\pi}{T_p}\csc\left(\frac{\pi T_e}{T_p}\right)\left(\cos\left(\frac{\pi T_e}{T_p}\right) - e^{T_e(s-a)}\right)}{(s-a)^2 + \left(\frac{\pi}{T_p}\right)^2} \tag{8.3}$$

The Laplace transform of the second part of the LF-model (the return phase) is shown in (8.4), where u(t) is the unit step function.

$$H_2(s) = \mathcal{L}\left\{\left[-\frac{E_e}{\varepsilon T_a}\left(e^{-\varepsilon t} - e^{-\varepsilon(T_0-T_e)}\right)\right][u(t) - u(t - T_0 + T_e)]\right\} \tag{8.4}$$

$$= -\frac{E_e}{\varepsilon T_a}\frac{s\left(1 - e^{\varepsilon(T_e-T_0)}\right) - \varepsilon e^{\varepsilon(T_e-T_0)}\left(1 - e^{s(T_e-T_0)}\right)}{s(s+\varepsilon)} \tag{8.5}$$

Since the LF pulse is the sum of the open phase and the return phase, the complete transform for the LF-model is:

$$H(s) = H_1(s) + H_2(s) \tag{8.6}$$

Subsequently, we substitute s with $j\omega$ in equation (8.6) to obtain the corresponding frequency response, which is equivalent to the frequency spectrum of the model. The absolute value of $H(j\omega)$ yields the amplitude spectrum: $A_1(f)$ and $A_2(f)$ for the open phase (8.7) and the return phase (8.9), respectively. The phase spectrum is derived from the angle determined by the real and imaginary parts of the transform: $\Phi_1(f)$ and $\Phi_2(f)$ for the open phase (8.8) and return phase (8.10), respectively. Also $Re_1(f)$, $Im_1(f)$, $Re_2(f)$, $Im_2(f)$ are (8.11), (8.12), (8.13), (8.14) respectively.

$$A_1(f) = E_e \frac{\sqrt{Re_1(f)^2 + Im_1(f)^2}}{\sqrt{\left(\left(\frac{\pi}{T_p}\right)^2 - (2\pi f)^2 + a^2\right)^2 + (4a\pi f)^2}} \tag{8.7}$$

$$\Phi_1(f) = \text{atan2}(Im_1(f), Re_1(f)) - \text{atan2}\left(-4a\pi f, \left(\frac{\pi}{T_p}\right)^2 - (2\pi f)^2 + a^2\right) \tag{8.8}$$

$$A_2(f) = \frac{\sqrt{Re_2(f)^2 + Im_2(f)^2}}{\sqrt{\left(-(2\pi f)^2\right)^2 + (2\pi f\varepsilon)^2}} \tag{8.9}$$

$$\Phi_2(f) = \text{atan2}(Im_2, Re_2) - \text{atan2}\left(2\pi f\varepsilon, -(2\pi f)^2\right) \tag{8.10}$$

$$Re_1(f) = \frac{\pi}{T_p \sin\left(\frac{\pi T_e}{T_p}\right)}\left(\cos\left(\frac{\pi T_e}{T_p}\right) - e^{-aT_e}\cos\left(2\pi f T_e\right)\right) - a \tag{8.11}$$

$$Im_1(f) = 2\pi f - \frac{\pi e^{-aT_e}\sin\left(2\pi f T_e\right)}{T_p \sin\left(\frac{\pi T_e}{T_p}\right)} \tag{8.12}$$

$$Re_2(f) = \frac{-E_e e^{\varepsilon(T_e - T_0)}\left(\cos\left(2\pi f\left(T_e - T_0\right)\right) - 1\right)}{T_a} \tag{8.13}$$

$$Im_2(f) = \frac{-E_e}{\varepsilon T_a}\left(2\pi f - 2\pi f e^{\varepsilon(T_e - T_0)} + \varepsilon e^{\varepsilon(T_e - T_0)}\sin\left(2\pi f\left(T_e - T_0\right)\right)\right) \tag{8.14}$$

By employing the phasor arithmetic outlined in Chapter 2, we obtain the comprehensive LF model spectrum as illustrated in equations (8.15) and (8.16), with atan2 defined according to 2.13:

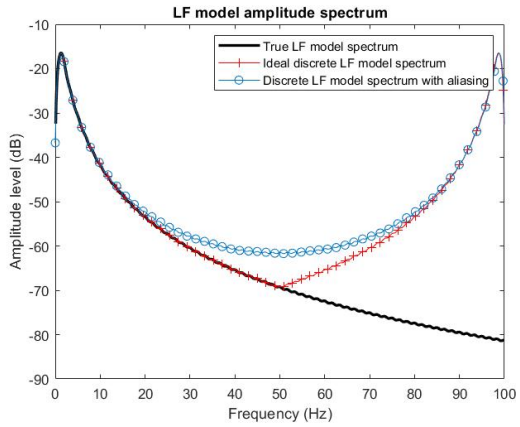
$$A = \sqrt{A_1^2 + A_2^2 + 2A_1A_2\cos\left(\Phi_1 - \Phi_2\right)} \tag{8.15}$$

$$\Phi = \text{atan2}\left(A_1\sin\Phi_1 + A_2\sin\Phi_2, A_1\cos\Phi_1 + A_2\cos\Phi_2\right) \tag{8.16}$$

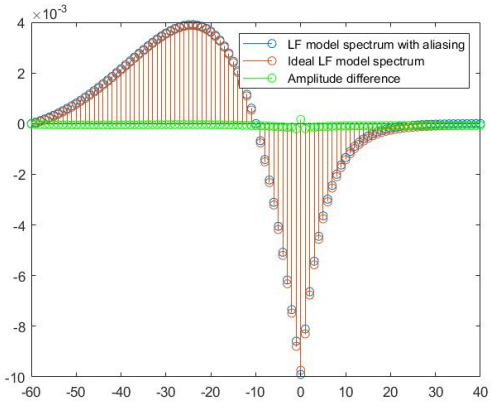
To shift the pulse to start at time-point $t_0 = 0$ rather than at T_e, a phase delay of $2\pi f T_e$ should be added to (8.16), so now Φ is (8.17):

$$\Phi = \text{atan2}\left(A_1\sin\Phi_1 + A_2\sin\Phi_2, A_1\cos\Phi_1 + A_2\cos\Phi_2\right) - 2\pi f T_e \tag{8.17}$$

After evaluating them and composing the signal and applying IFFT on matlab we get the following results.



(a) Amplitude spectrum of the LF-model showing the true spectrum (black), ideal discrete spectrum (red) and aliased discrete spectrum (blue). $F_s = 100Hz$



(b) LF pulse using the proposed method (red), compared to the corresponding sampled LF pulse (blue). The difference between the two is shown by the green line.

Chapter 9

The KLGLOTT88 model in the frequency domain

The KLGLOTT88 model, derived from the Rosenberg B model, has been employed in the Klatt synthesizer and various studies. This glottal waveform, denoted as U_g, is characterized by four parameters: the fundamental frequency F_0 (inverse of the period T_0), the amplitude of voicing A_V, the open quotient O_q, and the attenuation of a spectral tilt filter TL. When TL is set to 0 dB, the KLGLOTT88 model is identical to the Rosenberg B model. The model's equation and its derivative are modified according to the chosen parameters, influencing the shape of the glottal waveform.

For the modified KLGLOTT88 model, the Fourier transform was computed using equation (9.1), and the result is expressed by equation (9.2), and its written using $Re(f)$(9.3) and $Im(f)$(9.4) because of the length of the equation.

$$U_g'(t) = \begin{cases} 2ge^{-at}t - 3be^{-at}t^2, & 0 \le t \le O_qT_0 \\ \\ 0, & O_qT_0 \le t \le T_0 \end{cases} \tag{9.1}$$

$$H(f) = \frac{Re(f) + jIm(f)}{(a + j2\pi f)^3} \tag{9.2}$$

$$\begin{aligned} Re(f) =& 2ag - 2age^{-aT_0}\cos(2\pi fT_0) - 2g2\pi fe^{-aT_0}\sin(2\pi fT_0) \\ & - 2a^2gT_0e^{-aT_0}\cos(2\pi fT_0) + 2(2\pi f)^2gT_0e^{-aT_0}\cos(2\pi fT_0) \\ & - 4a2\pi fgT_0e^{-aT_0}\sin(2\pi fT_0) + 6be^{-aT_0}\cos(2\pi fT_0) - 6b \\ & + 6abT_0e^{-aT_0}\cos(2\pi fT_0) + 6b2\pi fT_0e^{-aT_0}\sin(2\pi fT_0) \\ & + 3a^2bT_0^2e^{-aT_0}\cos(2\pi fT_0) - 3b(2\pi f)^2T_0^2e^{-aT_0}\cos(2\pi fT_0) \\ & + 6ab2\pi fT_0^2e^{-aT_0}\sin(2\pi fT_0). \end{aligned} \tag{9.3}$$

$$\begin{aligned} Im(f) =& -6be^{-aT_0}\sin(2\pi fT_0) + 2g2\pi f - 2g2\pi fe^{-aT_0}\cos(2\pi fT_0) \\ & + 2age^{-aT_0}\sin(2\pi fT_0) - 4ag2\pi fT_0e^{-aT_0}\cos(2\pi fT_0) \\ & + 2a^2gT_0e^{-aT_0}\sin(2\pi fT_0) - 2g(2\pi f)^2T_0e^{-aT_0}\sin(2\pi fT_0) \\ & + 6b2\pi fT_0e^{-aT_0}\cos(2\pi fT_0) - 6abT_0e^{-aT_0}\sin(2\pi fT_0) \\ & + 6ab2\pi fT_0^2e^{-aT_0}\cos(2\pi fT_0) - 3a^2bT_0^2e^{-aT_0}\sin(2\pi fT_0) \\ & + 3b(2\pi f)^2T_0^2e^{-aT_0}\sin(2\pi fT_0). \end{aligned} \tag{9.4}$$

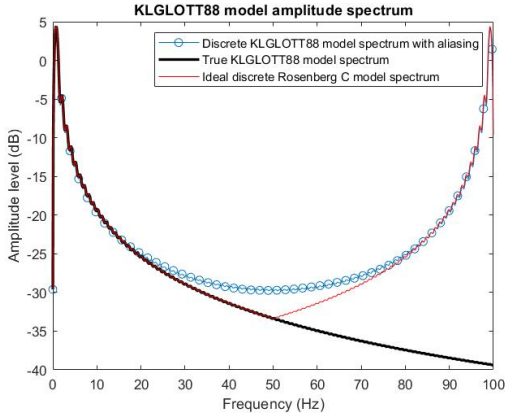
At last we have the equation that gives us the ideal model spectrum (9.5) and the equation that gives us the phase (9.6). Because this model has an abrupt closure there is only one equation, so there is no need for 2.11 and 2.12.

$$A = \frac{\sqrt{Re(f)^2 + Im(f)^2}}{\sqrt{\left(a^3 - 3a\left(2\pi f\right)^2\right)^2 + \left(3a^2\left(2\pi f\right) - \left(2\pi f\right)^3\right)^2}} \tag{9.5}$$

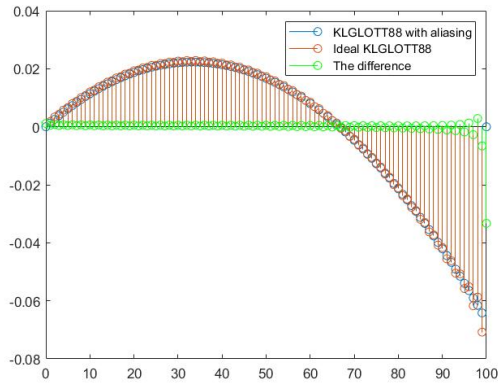
$$\Phi = \operatorname{atan2}\left(Im(f), Re(f)\right) - \operatorname{atan2}\left(3a^2 2\pi f - \left(2\pi f\right)^3, a^3 - 3a\left(2\pi f\right)^2\right) \tag{9.6}$$

After evaluating them and composing the signal and applying IFFT on matlab we get the following results.

(As mentioned previously, the KLGLOTT88 model was modified to make the method applicable to it.)



(a) Amplitude spectrum of the KLGLOTT88 model showing the true spectrum (black), ideal discrete spectrum (red) and aliased discrete spectrum (blue). $F_s = 100Hz$



(b) KLGLOTT88 pulse using the proposed method (red), compared to the corresponding sampled KLGLOTT88 pulse (blue). The difference between the two is shown by the green line.

Chapter 10

The R++ model in the frequency domain

For the modified first part of the R++ model, the Fourier transform was computed using equation (10.1), and the result is expressed by equation (10.2), and its written using $Re_1(f)$(10.11) and $Im_1(f)$(10.12) because of the length of the equation. For the second part of the R++ model, the Laplace transform had to be taken and its shown in (10.3), where u(t) is the unit step function as always.

$$U'_g(t) = \begin{cases} 4Ke^{-a(t+T_e)}(t+T_e)(T_p-t-T_e)(T_x-t-T_e), & -T_e \le t \le 0 \\ U'_g(0)\left(\dfrac{e^{-\frac{t}{T_a}} - e^{-\frac{T_0-T_e}{T_a}}}{1-e^{-\frac{T_0-T_e}{T_a}}}\right), & 0 \le t \le T_0 - T_e \end{cases} \tag{10.1}$$

As stated before for the first part of the model the first step of taking the Laplace transform is skipped because the Fourier transform was taken in the first place. The resulting amplitude spectrum is denoted as $A_1(f)$ for the open phase (10.7) and the phase spectrum is obtained by determining the angle formed by the real and imaginary parts of the transform, designated as $\Phi_1(f)$ for the return phase (10.8).

$$H_1(f) = \frac{Re_1(f) + jIm_1(f)}{(a+j2\pi f)^3} \tag{10.2}$$

$$H_2(s) = \mathcal{L}\left\{\left[U'_g(0)\left(\frac{e^{-\frac{t}{T_a}} - e^{-\frac{T_0-T_e}{T_a}}}{1-e^{-\frac{T_0-T_e}{T_a}}}\right)\right][u(t) - u(t-T_0+T_e)]\right\} \tag{10.3}$$

$$= \frac{U'_g(0)}{1-e^{-\frac{T_0-T_e}{T_a}}}\left[\frac{sT_a\left(1-e^{-\frac{T_0-T_e}{T_a}}\right) + e^{-\frac{T_0-T_e}{T_a}}\left(e^{-s(T_0-T_e)}-1\right)}{s\left(sT_a+1\right)}\right] \tag{10.4}$$

Following this, the transformation of s to $j\omega$ in equation (10.4) allows for the extraction of the frequency response, representing the frequency spectrum of the model. The resulting amplitude spectrum is denoted as $A_2(f)$ for the return phase (10.9). Additionally, the phase spectrum is obtained by determining the angle formed by the real and imaginary parts of the transform, designated as $\Phi_2(f)$ for the return phase (10.10).

$$H_2(f) = \frac{Re_2(f) + jIm_2(f)}{-(2\pi f)^2 T_a + j2\pi f} \tag{10.5}$$

The R++ pulse is the sum of the open phase and the return phase, the complete transform for the R++ model is:

$$H(f) = H_1(f) + H_2(f) \tag{10.6}$$

$$A_1(f) = \frac{\sqrt{Re_1(f)^2 + Im_1(f)^2}}{\sqrt{\left(a^4 - 6a^2(2\pi f)^2 + (2\pi f)^4\right)^2 + \left(4a^3 2\pi f - 4a(2\pi f)^3\right)^2}} \tag{10.7}$$

$$\Phi_1(f) = \operatorname{atan2}(Im_1(f), Re_1(f)) - \operatorname{atan2}\left(4a^3 2\pi f - 4a(2\pi f)^3, a^4 - 6a^2(2\pi f)^2 + (2\pi f)^4\right) \tag{10.8}$$

$$A_2(f) = \frac{\sqrt{Re_2(f)^2 + Im_2(f)^2}}{\sqrt{\left(-\left(2\pi f\right)^2 T_a\right)^2 + (2\pi f)^2}} \tag{10.9}$$

$$\Phi_2(f) = \text{atan2}(Im_2(f), Re_2(f)) - \text{atan2}\left(2\pi f, -\left(2\pi f\right)^2 T_a\right) \tag{10.10}$$

$$\begin{aligned}
Re_1(f) = 4K\ \big[&6\cos(2\pi f T_e) - 6e^{-aT_e} - 6aT_e\cos(2\pi f T_e) + 6(2\pi f T_e)\sin(2\pi f T_e) + \\
&+ 3a^2T_e^2\cos(2\pi f T_e) - 6a(2\pi f T_e^2)\sin(2\pi f T_e) - \\
&- 3(2\pi f)^2T_e^2\cos(2\pi f T_e) - T_e^3a^3\cos(2\pi f T_e) + T_e^3 3a(2\pi f)^2\cos(2\pi f T_e) + \\
&+ 3a^2(2\pi f)T_e^3\sin(2\pi f T_e) - (2\pi f)^3T_e^3\sin(2\pi f T_e) + \\
&+ (3T_e - T_p - T_x)\left(2a\cos(2\pi f T_e) - 2(2\pi f)\sin(2\pi f T_e) - 2ae^{-aT_e} - \right. \\
&-2a^2T_e\cos(2\pi f T_e) + 4a(2\pi f T_e)\sin(2\pi f T_e) + 2(2\pi f)^2T_e\cos(2\pi f T_e) + \\
&+a^3T_e^2\cos(2\pi f T_e) - 3a^2(2\pi f)T_e^2\sin(2\pi f T_e) - \\
&\left.-3a(2\pi f)^2T_e^2\cos(2\pi f T_e) + (2\pi f)^3T_e^2\sin(2\pi f T_e) + \right) \\
&+ (3T_e^2 - 2T_eT_p + T_pT_x - 2T_eT_x)\left(a^2\cos(2\pi f T_e) - 2a(2\pi f)\sin(2\pi f T_e) - \right. \\
&-(2\pi f)^2\cos(2\pi f T_e) - a^2e^{-aT_e} + (2\pi f)^2e^{-aT_e} - \\
&-a^3T_e\cos(2\pi f T_e) + 3a^2(2\pi f)T_e\sin(2\pi f T_e) + \\
&\left.+3a(2\pi f)^2T_e\cos(2\pi f T_e) - (2\pi f)^3T_e\sin(2\pi f T_e)\right) + \\
&+ (T_eT_pT_x - T_e^2T_p - T_e^2T_x + T_e^3)\left(a^3\cos(2\pi f T_e) - 3a^2(2\pi f)\sin(2\pi f T_e) - \right. \\
&\left.-3a(2\pi f)^2\cos(2\pi f T_e) + (2\pi f)^3\sin(2\pi f T_e) - a^3e^{-aT_e} + 3a(2\pi f)^2e^{-aT_e}\right]
\end{aligned} \tag{10.11}$$

$$\begin{aligned}
Im_1(f) = 4K\ \big[&6\sin(2\pi f T_e) - 6aT_e\sin(2\pi f T_e) - 6(2\pi f)T_e\cos(2\pi f T_e) + 3a^2T_e^2\sin(2\pi f T_e) + \\
&+ 6a(2\pi f)T_e^2\cos(2\pi f T_e) - 3(2\pi f)^2T_e^2\sin(2\pi f T_e) - 3a^2(2\pi f)T_e^3\cos(2\pi f T_e) + \\
&+ (2\pi f)^3T_e^3\cos(2\pi f T_e) - a^3T_e^3\sin(2\pi f T_e) + 3a(2\pi f)^2T_e^3\sin(2\pi f T_e) + \\
&+ (3T_e - T_p - T_x)\left(2a\sin(2\pi f T_e) + 2(2\pi f)\cos(2\pi f T_e) - 2(2\pi f)e^{-aT_e} - \right. \\
&- 2a^2T_e\sin(2\pi f T_e) - 4a(2\pi f)T_e\cos(2\pi f T_e) + 2(2\pi f)^2T_e\sin(2\pi f T_e) + a^3T_e^2\sin(2\pi f T_e) + \\
&\left.+ 3a^2(2\pi f)T_e^2\cos(2\pi f T_e) - 3a(2\pi f)^2T_e^2\sin(2\pi f T_e) - (2\pi f)^3T_e^2\cos(2\pi f T_e)\ \right) + \\
&+ \left(3T_e^2 - 2T_eT_p + T_pT_x - 2T_eT_x\right)\left(a^2\sin(2\pi f T_e) + 2a(2\pi f)\cos(2\pi f T_e) - \right. \\
&- (2\pi f)^2\sin(2\pi f T_e) - 2a(2\pi f)e^{-aT_e} - a^3T_e\sin(2\pi f T_e) - 3a^2(2\pi f)T_e\cos(2\pi f T_e) + \\
&\left.+ 3a(2\pi f)^2T_e\sin(2\pi f T_e) + (2\pi f)^3T_e\cos(2\pi f T_e)\ \right) + \\
&+ \left(T_eT_pT_x - T_e^2T_p - T_e^2T_x + T_e^3\right)\left(a^3\sin(2\pi f T_e) + 3a^2(2\pi f)\cos(2\pi f T_e) - \right. \\
&\left.- 3a(2\pi f)^2\sin(2\pi f T_e) - (2\pi f)^3\cos(2\pi f T_e) - 3a^2(2\pi f)e^{-aT_e} + (2\pi f)^3e^{-aT_e}\ \right]
\end{aligned} \tag{10.12}$$

$$Re_2(f) = \frac{U_g'(0)}{1 - e^{-\frac{T_0 - T_e}{Ta}}}\left[e^{-\frac{T_0 - T_e}{Ta}}\left(\cos\left(2\pi f(T_0 - T_e)\right) - 1\right)\right] \tag{10.13}$$

$$Im_2(f) = \frac{U_g'(0)}{1 - e^{-\frac{T_0 - T_e}{Ta}}}\left[2\pi f Ta - e^{-\frac{T_0 - T_e}{Ta}}\left(2\pi f Ta + \sin\left(2\pi f(T_0 - T_e)\right)\right)\right] \tag{10.14}$$

By employing the phasor arithmetic outlined in Chapter 2, we obtain the comprehensive R++ model spectrum as illustrated in equations (10.15) and (10.16), with atan2 defined according to 2.13:

$$A = \sqrt{A_1^2 + A_2^2 + 2A_1A_2\cos\left(\Phi_1 - \Phi_2\right)} \tag{10.15}$$

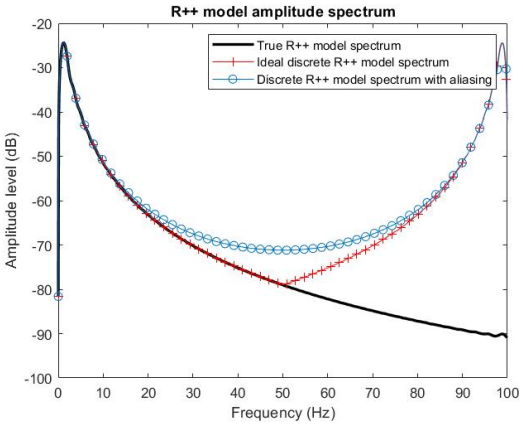
$$\Phi = \text{atan2}\left(A_1\sin\Phi_1 + A_2\sin\Phi_2, A_1\cos\Phi_1 + A_2\cos\Phi_2\right) \tag{10.16}$$

To shift the pulse to start at time-point $t_0 = 0$ rather than at T_e, a phase delay of $2\pi f T_e$ should be added to (10.16), so now Φ is (10.17):

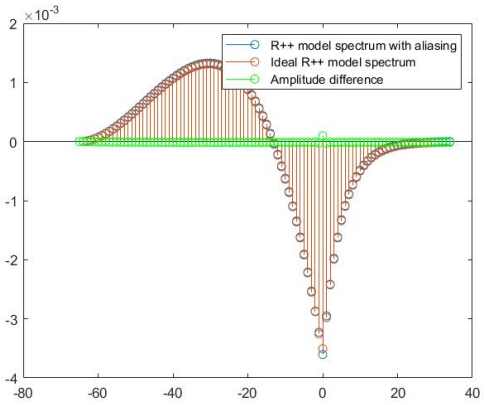
$$\Phi = \text{atan2}\left(A_1\sin\Phi_1 + A_2\sin\Phi_2, A_1\cos\Phi_1 + A_2\cos\Phi_2\right) - 2\pi f T_e \tag{10.17}$$

After evaluating them and composing the signal and applying IFFT on matlab we get the following results.

(As mentioned previously, the R++ model was modified to make the method applicable to it.)



(a) Amplitude spectrum of the R++ model showing the true spectrum (black), ideal discrete spectrum (red) and aliased discrete spectrum (blue). $F_s = 100Hz$



(b) R++ pulse using the proposed method (red), compared to the corresponding sampled R++ pulse (blue). The difference between the two is shown by the green line.

Chapter 11

Additional plots

Below, additional plots are presented, demonstrating the stability and versatility of the method when applied to signals with varying fundamental frequencies and parameter values. Two plots are dedicated to the Rosenberg C model, two plots correspond to the LF model, two plots to the KLGLOTT88 model and two plots to R++ model with specific parameter values.

Rosenberg C variables (1)	**Rosenberg C variables (2)**
$A = 0.6381$	$A = 0.6427$
$T_0 = 0.004s$	$T_0 = 0.0167s$
$T_p = 0.0024s$	$T_p = 0.0095s$
$T_n = 0.0016s$	$T_n = 0.0071s$
$a \to 0$	$a \to 0$
$F_s = 20000Hz$	$F_s = 5000Hz$

Table 11.1: The additional Rosenberg C variable's table

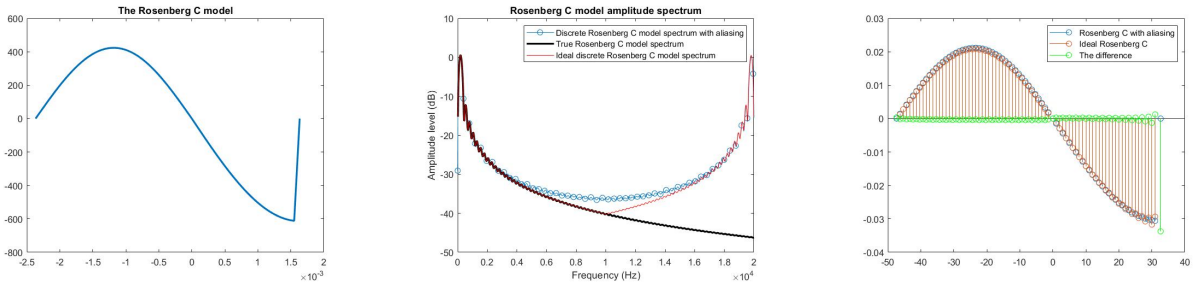


Figure 11.1: Rosenberg C variables (1)

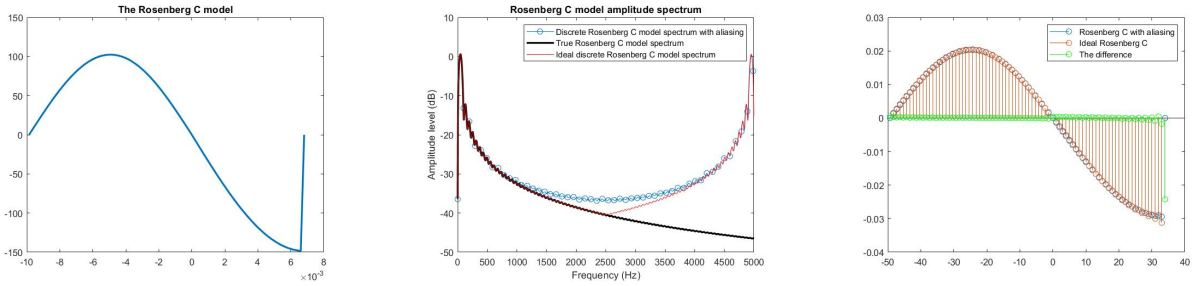


Figure 11.2: Rosenberg C variables (2)

LF model variables (1)	**LF model variables (2)**
$E_e = 1$	$E_e = 1$
$T_0 = 0.004s$	$T_0 = 0.0167s$
$T_e = 0.003s$	$T_e = 0.0128s$
$T_p = 0.0027s$	$T_p = 0.0111s$
$T_b = 0.001s$	$T_b = 0.0038s$
$a = 1500$	$a = 300$
$e = 8000$	$e = 1500$
$F_s = 20000Hz$	$F_s = 5000Hz$

Table 11.2: The additional LF model variable's table

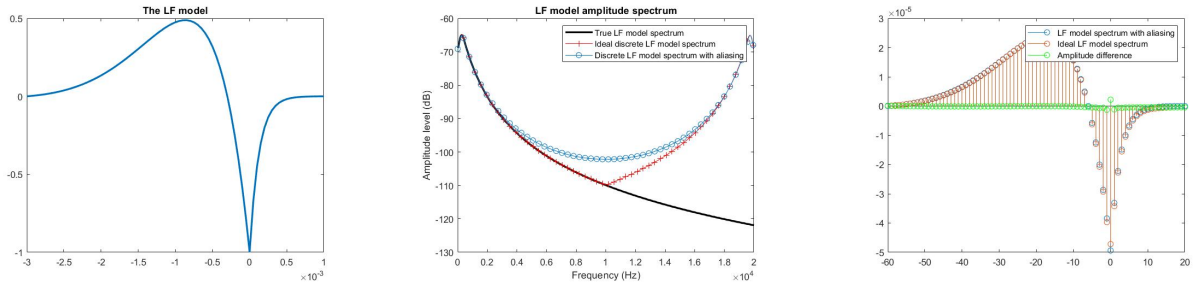


Figure 11.3: LF model variables (1)

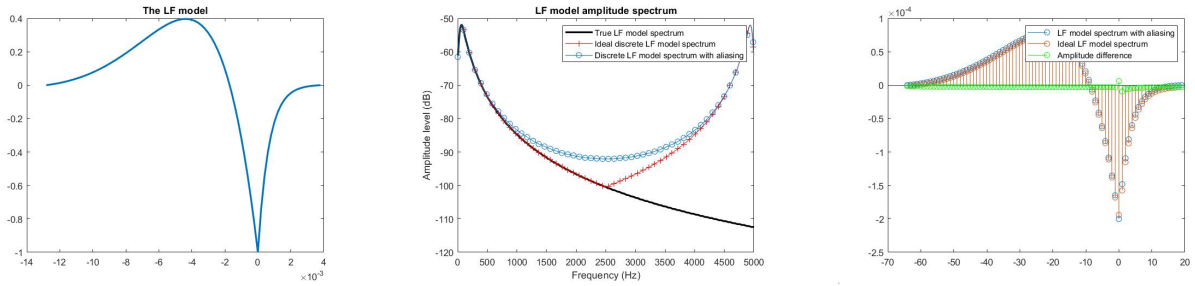


Figure 11.4: LF model variables (2)

KLGLOTT88 model variables (1)	**KLGLOTT88 model variables (2)**
$T_0 = 0.004s$	$T_0 = 0.0167s$
$a \to 0$	$a \to 0$
$g = 1687.5$	$g = 405$
$b = 421875$	$b = 24300$
$F_s = 20000Hz$	$F_s = 5000Hz$

Table 11.3: The additional KLGLOTT88 model variable's table

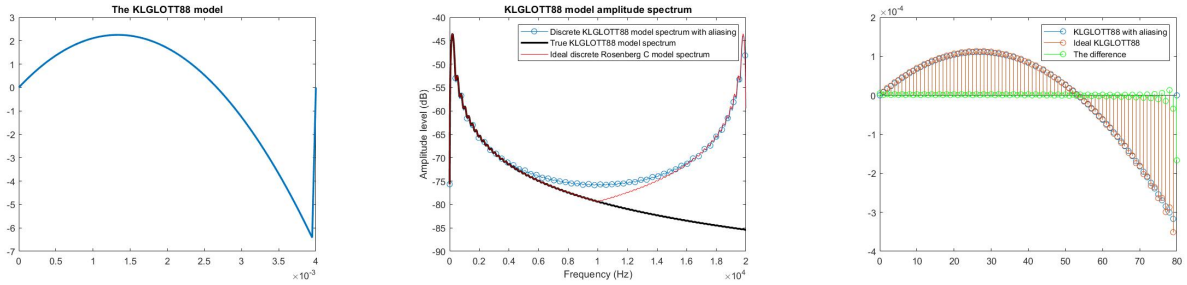


Figure 11.5: KLGLOTT88 model variables (1)

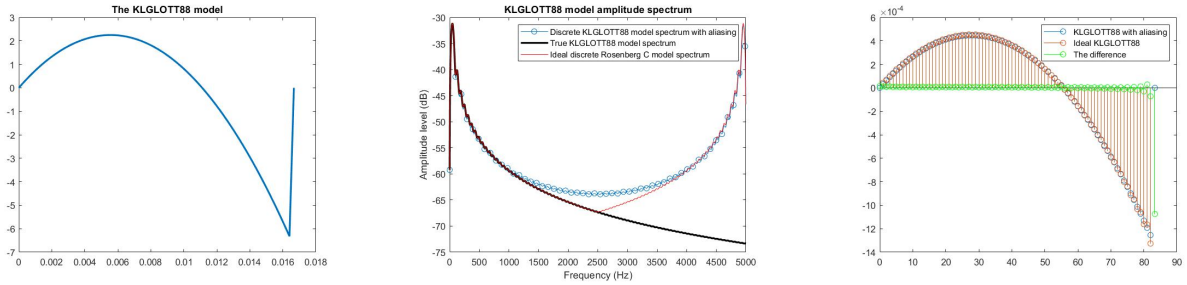


Figure 11.6: KLGLOTT88 model variables (2)

R++ model variables (1)	**R++ model variables (2)**
$T_0 = 0.004s$	$T_0 = 0.0167s$
$a \to 0$	$a \to 0$
$T_e = 0.0026s$	$T_e = 0.0108s$
$T_p = 0.0021s$	$T_p = 0.0086s$
$T_a = 0.0002$	$T_a = 0.001$
$T_x = 0.000010046$	$T_x = -0.00062681$
$U_g'(0) = -0.000000023062$	$U_g'(0) = -0.0000017716$
$am = 0.7971$	$am = 0.7971$
$D = 0.9936$	$D = 0.9829$
$K = -1.6228$	$K = -1.6228$
$F_s = 20000Hz$	$F_s = 5000Hz$

Table 11.4: The additional R++ model variable's table

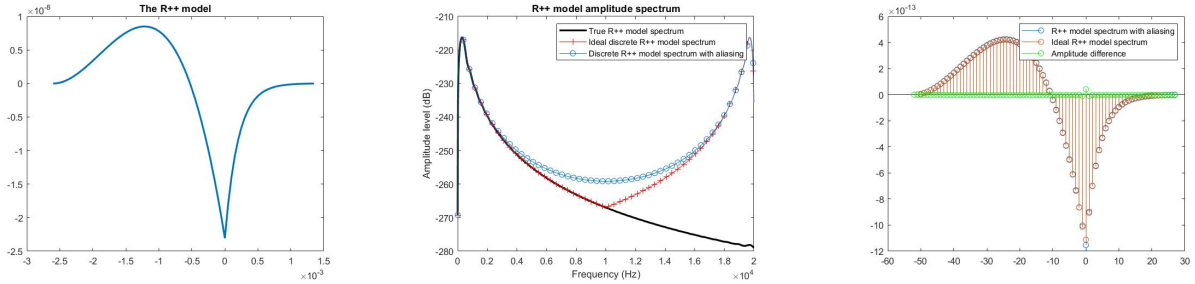


Figure 11.7: R++ model variables (1)

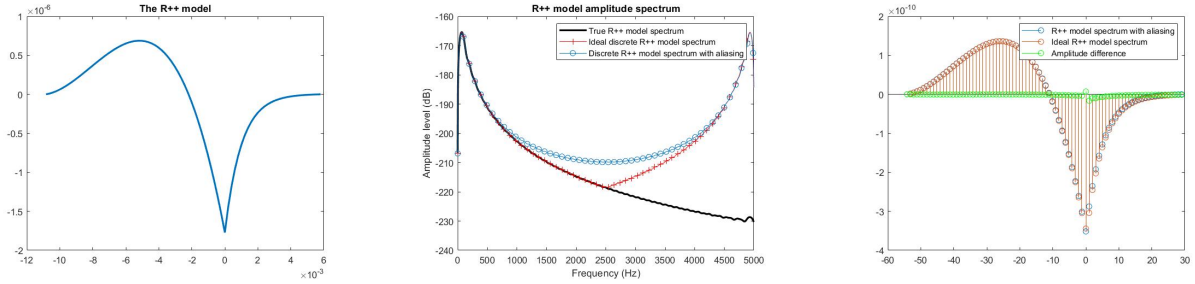


Figure 11.8: R++ model variables (2)

Chapter 12

Conclusions

The source-filter theory serves as a fundamental framework for understanding human speech production. It explains how the vocal folds generate the source, which can be influenced by both oscillations and turbulent airflows. The vocal tract acts as a filter, shaping the spectral structure of the source sounds through resonances. Maintaining independence between the source and the filter is crucial for effective language-based communication, allowing for the flexible articulation of various phonemes. Resonance tuning, observed in soprano singers and certain animals, exemplifies the practical application of this theory. Furthermore, the interaction between the source and filter can lead to voice instabilities, particularly when the fundamental frequency crosses vocal tract resonances, as seen in singing scenarios. This interaction highlights the complex dynamics involved in speech production and its susceptibility to external influences.

The method introduced allows for the implementation of time domain glottal flow models without encountering aliasing distortion. A crucial aspect of this method involves creating a frequency domain representation of the time domain model. In this context, such a representation is outlined for the Rosenberg C, LF, KL-GLOTT88 and R++ glottal flow models. For each of these models, closed-form expressions for the amplitude and phase spectrum are provided, facilitating the direct computation of the spectrum based on the respective model's parameters.

The stability of this method is demonstrated through consistent and reproducible results across the four different glottal flow models discussed in this thesis, even when varying parameters or fundamental frequencies. Chapter 11 contains additional plots that showcase the method's robustness and reliability under different conditions, affirming its consistency in providing accurate results across various scenarios and model configurations.

The modifications applied to these models are minimal in terms of the degree of change, as the term e^{-at} approaches 1 when a tends towards 0. This suggests that the adjustments made to the models have a limited impact and are primarily influenced by the behavior of the exponential term under the specified conditions.

In comparison to alternative methods, this approach entirely eliminates aliasing distortion, eliminates the need for specific anti-aliasing functions, and eliminates the necessity for post-filtering to rectify spectral amplitude levels. Moreover, it is conceptually intuitive, and once the frequency domain representation is acquired, implementation is relatively straightforward.

Bibliography

[1] Christer Gobl (2021), *"The LF Model in the Frequency Domain for Glottal Airflow Modelling without Aliasing Distortion"*, Phonetics and Speech Laboratory, School of Linguistic, Speech and Communication Sciences Trinity College Dublin, Ireland.

[2] Boris Doval, Christophe d'Alessandro (2006), *"The Spectrum of Glottal Flow Models"*, ACTA ACUSTICA UNITED WITH ACUSTICA Vol. 92 (2006) 1 – 1.

[3] Gunnar Fant, Johan Liljencrants, and Qi-guang Lin (1985), *"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] Tahir Mushtaq QURESHI, Khalid Saifullah SYED (2011), *"A New Approach to Parametric Modeling of Glottal Flow"*, Bahauddin Zakariya University Center for Advance Studies in Pure and Applied Mathematics Bosan Road, 60800, Multan, Pakistan.

[5] E. Cataldo, D. Bahiano (2021), *"Stochastic models of glottal pulses from the Rosenberg and Liljencrants-Fant models with unified parameters"*, aUniversidade Federal Fluminense, Telecommunications Department and Graduate program in Telecommunications Engineering, R. Passo da P´atria, 156/209 - S˜ao Domingos, Niter´oi, RJ - Brazil - 24210-240.

[6] Dennis H. Klatt and Laura C. Klatt (1990), *"Analysis, synthesis, and perception of voice quality variations among female and male talkers"*, Research Laboratory of Electronics Room 36-523, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

[7] Christophe d'Alessandro, Baris Bozkurt, Boris Doval, Thierry Dutoit, Nathalie Henrich, Vu Ngoc Tuan, Nicolas Sturmel (2007), *"Phase-based methods for voice source analysis"*, LIMSI-CNRS Orsay, France, Izmir Institute of Technology, Izmir, Turkey, TCTS-FPMs, Mons, Belgium, DPC-GIPSA-Lab Grenoble

[8] A. E. Rosenberg (1971), *"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

[9] George P. Kafentzis (2008), *"On the glottal flow derivative waveform and its properties"*, A time/frequency study, COMPUTER SCIENCE DEPARTMENT UNIVERSITY OF CRETE

[10] Christopher J Bleakley (1995), *"AN INVESTIGATION INTO GLOTTAL WAVEFORM BASED SPEECH CODING"*, School of Electronic Engineering Dublin City University Dublin 9, Ireland

[11] Gunnar Fant Qi-guang Lin (1988), *"Frequency domain interpretation and derivation of glottal flow parameters"*, STL-QPSR vol. 29, no.2-3, pp. 001-021, 1988

[12] OPAZO CAMPUSANO, BENJAMÍN (2022), *"Voice Quality Modification using the WORLD Vocoder"*, UNIVERSIDAD TECNICA FEDERICO SANTA MARIA DEPARTAMENTO DE ELECTRONICA

[13] Isao Tokuda (2021), *"The Source–Filter Theory of Speech"*, Phonetics/Phonology

[14] The source-filter model of speech production , Linguistic Phonetics 24.915/24.963

[15] Speech production and acoustic properties , Tom Bäckström 2021

[16] K. Ishizaka, J. L. Flanagan (1972)*"Synthesis of Voiced Sounds From a Two-Mass Model of the Vocal Cords"*, The Bell System Technical Journal

[17] Johan Liljencrants (1991), *"A translating and rotating mass model of the vocal folds"*, STL-QPSR 1, 1-18.

[18] Pelorson, X (1994), *"Theoreticaland experimental study of quasisteady-flow separation within the glottis during phonation.Application to a modified two-mass model"*, J. Acoust. Soc. Am. 96, 3416-3431.

[19] HANSPETER HERZEL, CARSTEN KNUDSEN (1995), *"Bifurcations in a Vocal Fold Model"*, Nonlinear Dynamics 7: 53-64, 1995

[20] Brad H. Story, Ingo R. Titze (1995), *"Voice simulation with a body-cover model of the vocal folds"*, The Journal of the Acoustical Society of America 97(2):1249-60

[21] David A Berry, Ingo R Titze (1996), *"The normal modes of vocal fold tissues"*, The Journal of the Acoustical Society of America, 2656-2656, Acoustical Society of America

[22] de Vries M.P., Schutte H.K., Verkerke G.J. (1999), *"Determination of parameters for lumped parameter model of the vocal fold using a finite-element method approach"*, Journal of the Acoustical Society of America, 106, 6, 3620-3628.

[23] Gunter H.E. (2003), *"A mechanical model of vocal-fold collision with high spatial and temporal resolution"*, Journal of the Acoustical Society of America, 113, 2, 994-1000.