Computational Model and Simulation of Articulatory Mechanism in *Yoruba* Voiced Speech

ADEGBITE Adewuyı Adetayo *adewuyi.adegbite@aaua.edu.ng*
Department of Computer Science,
Adekunle Ajasin University, Akungba-Akoko, Nigeria

ODEJOBI, Ajadı Odetunjı *oodejobi@yahoo.com*
Department of Computer Science and Engineering,
Obafemi Awolowo. University, Ile-Ife, Nigeria

LAYENI, Olawanle P. *olawanle.layeni@gmail.com*
Department of Mathematics,
Obafemi Awolowo. University, Ile-Ife, Nigeria

Abstract

This study examined the physical, electrical and mathematical models used in the dynamics of voiced sound production. It formulated and designed a computational model for the standard *Yoruba* voiced sounds, and the designed model also implemented. This was with a view to developing a *Yoruba* speech recognition and text-to-speech model. The mechanism of human speech production articulatory process documented in the existing literature was examined and analysed. The mechanical coupling in the vocal cords of the oral and that of the nasal cavity were studied. Then a computational model, which defined real variables over Standard *Yoruba* voice speech production process was formulated and designed using algorithm. The design was then implemented using an appropriate numerical computational tool called Matlab. The implemented novel model established that more volume in time of air will be needed for nasal vowels than oral vowels in the production of *Yoruba* voiced speech. A minimum of $238cm^3$ of air is needed for nasal cavity while a maximum of $171cm^3$ is needed for oral cavity. It was deduced that the change in the damping coefficient along the vocal cord does not affect the response rate of the speech organ whose rise time remains 0.9ms while a change in the spring constant causes changes in the response rate parameters. Whenever damping coefficient is constant, that is, either 0.1 or 0.2 over the points positioned with masses, even if the vocal chords of individuals have different assigned mass values, the speech production is still the same and observed as normal. The study concluded by establishing a computational model for nasalized *Yoruba* vowels. This model has the utility of serving as a resource for *Yoruba* speech recognition and text-to-speech application.

Keywords: Yoruba Vowels, Speech Production Process, Oral Cavity, Nasal Cavity, Articulatory Mechanism.

1 INTRODUCTION

To have a full grip and understanding of articulatory speech synthesis, there is need to understand the natural human speech production process as it will be very hard to imitate anything without first understanding the object of imitation [24]. For a long time, researchers have been trying to simulate the human speech production process using various means [41] including direct speech waveform, spectral waveform [43], glottal excitation models [42], etc.

According to [13, 37], the process of voiced sounds production can be described as air coming from the lungs forced through the narrow space between the two vocal folds, which are set in motion at a frequency governed by the tension and contraction of their tissues and muscles around the lung cavity. The vocal folds change the continuous flow that comes from the lungs into a series of pulses, causing periodic vibration whose rate gives the pitch of the sound. The resulting periodic puffs of air act as an excitation input, or source, to the vocal tract. Then, as the flow passes through the oral and nasal cavities it is amplified and changed until it is finally radiated from the mouth and/or nose as the case may be.

Articulatory organs are the vocal tract which begins at the vocal cords and ends at the lips for oral sounds, vocal cord which is sometimes called as the larynx and the nasal tract begins at the velum and ends at the nostrils. The part of the acoustic effects of such actions and interactions that is radiated (mainly) through the lips and nostrils constitutes the speech signal [17, 37].

The glottal signals are intimately related to the anatomic and physiological characteristics of the larynx [7, 8]. There are always rapid changes in the shape of vocal tracts in sound production due to the movements of the various speech articulators such as tongue, lips, jaw or larynx which involves the muscles, nerves and brain working together to execute the voice production.

1.1 The Standard *Yoruba* Language

Yoruba is one of the four major languages in the continent of Africa. Other languages in these category are Arabic, Hausa, and Swahili. *Yoruba* language is a Niger-Congo language spoken majorly in the Western part of Africa with over 40 million speakers [2]. The standard *Yoruba* language has three phonological contrastive tones [3] : high, mid (which is the default tone), low. A high tone is represented with an acute accent mark ('), a mid tone is most times left unmarked but in certain situations marked with a macron (-) and a low tone is represented with a grave accent mark ('). The accent marks are placed on the vowels of each syllable in a word. Though there are various dialects of the *Yoruba* language which is based on the geographical location, the Standard *Yoruba* Language is the generally understood and collectively accepted in writing. The *Yoruba* alphabet consists of 25 letters, and uses the familiar Latin characters. The *Yoruba* letters are made up of eighteen (18) consonants and seven (7) vowels. The consonants are b, d, f, g, gb, h, j, k, l, m, n, p, r, s, *.s*, t, w, y. The consonant "gb" is a diagraph (that is a consonant represented with two letters).

The vowels in *Yoruba* can classified into two and they are oral and nasal vowels. Oral vowels are produced entirely and completely when air flows through the mouth. The chart in Figure 1 shows the top-to-bottom dimension representing the vowel height or openness, that is, the higher positions on the chart correspond to a higher position of the tongue in the mouth. The *[i]* and *[u]* produced the highest position of the tongue while *[a]* produced the lowest position of the tongue. The left-to-right dimension of the chart corresponds approximately to the front-to-back position of the tongue in the mouth. The three vowels *[c], [o],* and *[u]* are rounded which means they are produced with rounded lips while the remaining four vowels are unrounded.

A nasalized vowel is a vowel which in the process of the sound production, the soft palate is lowered opening the velum which allows air to go through both the oral (majorly the mouth) and the nasal cavities (majorly the nose). The position of each of the nasal vowels are approximately equivalent to that of the corresponding oral vowel in the chart below. The mid vowels are not nasalized. There are five nasalized vowels in the standard *Yoruba* language and they are *an*, *en*, *in*, *on*, and *un*.

Also, we have two syllabic nasals and they are *m*, *n*. Unlike some other languages, the *Yoruba* language does not have diphthongs, which makes each vowel to be pronounced as separate syllables [3].

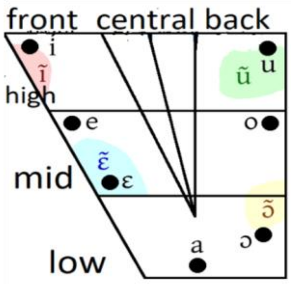


FIGURE 1: *Yoruba*Vowel Chart Source: [3].

1.2 Concatenative Speech Synthesis

In synthesizing speech, there are three(3) approaches to the synthesis and they are concatenative speech synthesis, formant speech synthesis and articulatory speech synthesis [45]. In concatenative speech synthesis, units of sounds are recorded which serves as segments of data. The knowledge-based database possess and acquire a lot of segments of data in the synthesizer which are embedded and chained up. The collection of data that are in the database segments are diphones, half-syllables, triphones etc. Since segments to be chained up have been extracted from different words and different phonetic context, there are sometimes amplitude mismatches and some audible discontinuities are sometimes also noticed. In concatenation, the type of segments chosen, the corpus they were extracted from, the segments quality, the amount of degradation introduced by the speech coding phase, the capabilities of the concatenation algorithm, the prosody matching efficiency are all the factors that decide the quality of speech produced [16, 44, 47]. The unit selection involves the definition of the inventory of units as well as selecting the appropriate unit for a given phonetic context.

An advantage of speech unit concatenation is that it is easy to produce realistic coarticulation effects, if suitable speech units are chosen. It is also appealing in terms of its simplicity, in that all knowledge concerning the synthetic message is inherent to the speech units to be concatenated.

1.3 Formant Speech Synthesis

The seemingly efficient representation of the speech signal in the spectral domain led to the development of formant synthesizers used in formants' speech synthesis. Formants are the acoustic product of a series of bandpass filters that outline the sound primarily produced by activities of the vocal folds in the larynx and emanated from the mouth. The modeling is done in the frequency domain by a set of resonators [46]. [25] describes formant synthesis as a parametric approach which applies a set of rules for controlling the frequencies and amplitudes of the formants and the characteristics of the pressure from the excitation source. For the formant synthesis to be used, there are input parameters which must be tracked, the largeness of these parameters is one of the fundamental limit of the formant speech model. The filtering process takes place in the entire vocal tract, which consists of the pharynx, as well as the laryngeal, nasal, and oral cavities. The cross-sectional diameters and length of the vocal tract determines the location of the formants' acoustic energy. Formants are gotten using rule based approach from mathematical waveforms and these are secondary data as seen in the work done by [40].

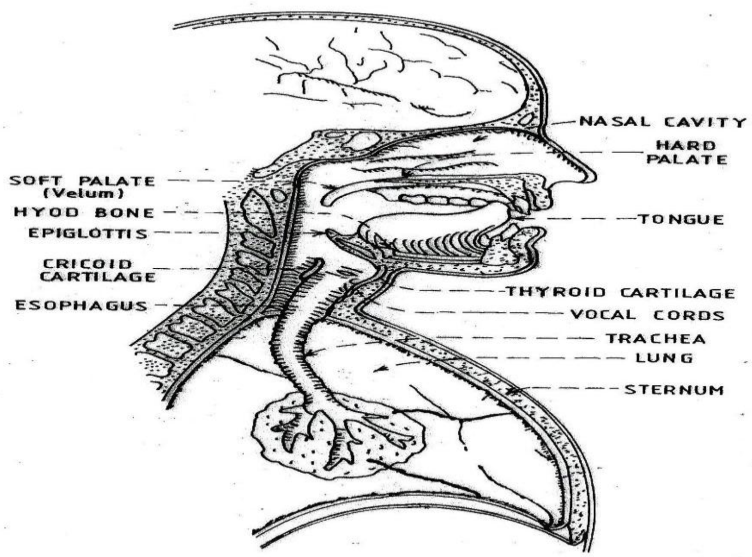


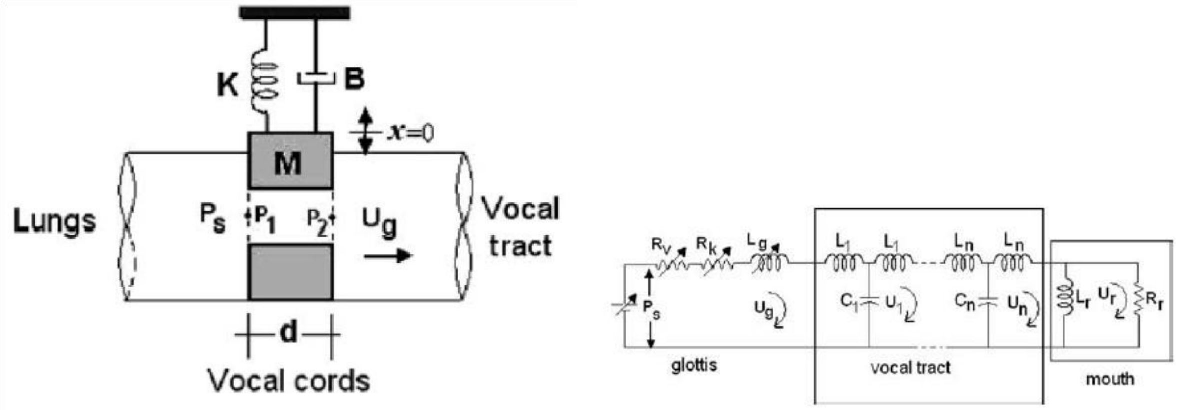
FIGURE 2: Articulatory organs in speech production Source: [20].

1.4 Articulatory Speech Synthesis

According to [16], Articulatory synthesis attempts the production of speech by understanding first of all how the vocal apparatus modulate in shape during the speech production and an aspect of these was done by [4] in examining the basic building blocks of vocal fold vibration. The speech organ system is modeled instead of its acoustic characteristics or its signal ([33, 23]). The acoustic problem of how these movements translate into sounds is also been critically studied. The input parameters used are the positions of the articulators studied and been modeled; and these parameters are specified through time. In these types of modeling, the articulatory organs itself are been modeled, Figure 2 shows the organs involved in the production of sounds. The vocal folds are soft tissue structures contained within the cartilaginous framework of the larynx. Their location in the neck and their ability to abduct (move apart) during respiration and to adduct (move together) during phonation makes the vocal folds the point of division between the subglottal and supraglottal airways [30]. In speech production mechanism, these speech organs are often categorised and broken down into three sections. The first section is the subglottal area comprising of the lungs, diaphragm, and trachea. The second section is the glottal area which is the area covered by the vocal cord while the third section is called the supraglottal area which is the area starting from the vocal tract extending to the resonant cavities between the glottis and the mouth and nose which is really the pharyngeal, the oral, and the nasal cavities.

In speech modeling, [10] postulated that the key differences in the various models are the manner in which the glottal waveform is generated and also the manner in which the subglottal area, glottal area and supraglottal area affect each other.

Articulatory synthesis has the potential to synthesize human-like speech sounds [16] and has more flexibility than conventional approaches. The speech produced is more accurate and sounds are authentic. The focus of these work is in the articulatory modeling of speech.



(a) One mass mechanical model

(b) One mass Electrical/ Acoustic circuits

FIGURE 3: One Mass System.

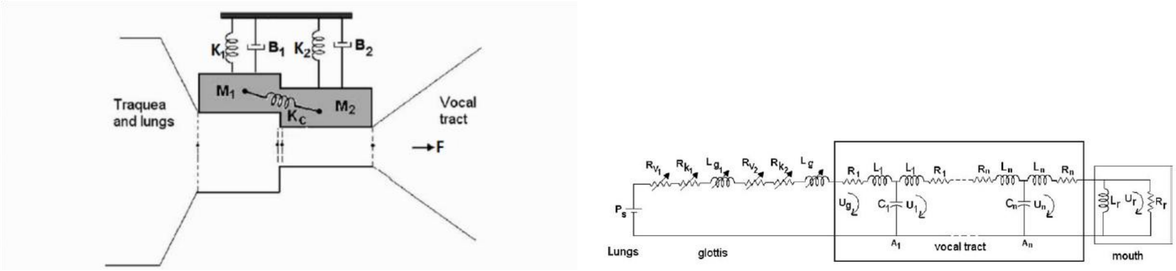
Production of speech in a language (i.e. speech = sound + language) can be better studied by examining articulatory or acoustic properties of the speech sounds since the voice quality is dependent on the vocal cord vibration [17]. According to [7], vocal cords play an important role in voice production.

Low-dimensional models of the vocal folds was discovered by [35] to oscillate in a semi-realistic way with muscle activations, but with a lot of rules necessary to capture the parameters. More research on articulatory organs especially the vocal cord were vividly studied by [17], the vocal cord been the main source of voiced-sound in speech was approximated as a two mass self-oscillating model. Though, the one mass model produced could approximate and simulate many of the properties but it was inadequate to produce the physiological details in vocal cord behavior. Figure 3(a) shows the mechanical model and its corresponding acoustic and electrical model is shown in Figure 3(b). So, for more physiological properties and details to be gotten, more mass are been added. According to [17], the two mass mechanical models as shown in Figure 4(a) are theoretically enough to give a true representation of the properties needed. The network or electrical model is gotten as seen in Figure 4(b) showing all the necessary properties in speech production which major on the oral cavity.

According to [7], for voice production articulatory system to be modeled, the various speech organs needs to be studied, and the various parts that had been considered are shown in Figure 5; and it could be seen that the parts studied were only the oral cavity. With the voice production system, some work has been done on them and mechanical models have been developed.

[7] were able to show the mathematical formula and mechanical model used which was adapted from [15]. The model has a mass M which is measured in Kilogramme (Kg), B is the constant of the damper in the one-mass model which is measured in Newton-second/metre (Ns/m), there is a forcing function F(t) which is measured in Newton(N) acts on the mass. The displacement x of the mass M is measured in metre. The first and second order derivative of the displacement with respect to time are represented as $x^{\cdot}$ and $x^{\cdot\cdot}$ respectively.

The mechanical model is a one-mass model and the mathematical formular gotten was:



(a) Two mass mechanical model

(b) Two mass Electrical/ Acoustic Circuits

FIGURE 4: Two Mass System.

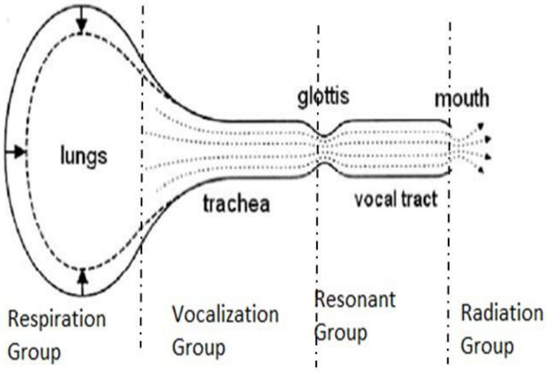


FIGURE 5: Distinct Speech Organ Group.

$$M\ddot{x} + B\dot{x} + Kx = F(t) \qquad (1)$$

In the two-mass model, it has two masses M_1 and M_2 which are measured in Kilogramme (Kg), B_1 and B_2 which is measured in Newton-second/metre (Ns/m) are the constant of the damper on mass M_1 and M_2 respectively, there is are two forcing function F_1 and F_2 which are measured in Newton(N) acts on the mass M_1 and M_2 respectively. The displacement x_1 and x_2 of the mass M_1 and M_2 respectively are measured in metre. The first and second order derivative of the displacement of the first mass with respect to time are represented as $\dot{x}_1$ and $\ddot{x}_1$ respectively while that of the second mass are represented as $\dot{x}_2$ and $\ddot{x}_2$ respectively.

When more mass were added to the one mass to make it two-mass model the mathematical model developed was

$$M_1\ddot{x}_1 + S_1(x_1) + B_1\dot{x}_1 + K_c(x_1 - x_2) = F_1 \qquad (2)$$

$$M_2\ddot{x}_2 + S_2(x_2) + B_2\dot{x}_2 + K_c(x_2 - x_1) = F_2 \qquad (3)$$

2 MODEL IMPLEMENTATION

The full understanding of production of sounds through the speech organs is needed for modeling of articulatory speech synthesis. In production of *Yoruba* sounds, the air that is been pumped through pressure from the lungs comes out through a narrow space between two vocal folds. The tension in the tissues of the speech organs changes continuously and these set the air flowing through it into motion of series of pulses. The air is then sent to the nasal and oral tract to be amplified through the nostril and the mouth as can been seen in Figure 6.

In the oral cavity, we have the lungs, trachea, vocal chords, vocal folds all ending with the mouth in the production of voice or sounds. Oral vowels are vowels without nasalization. In *Yoruba* language, the oral vowels are *a ,e ,e. ,i ,o ,o. ,u*.

Nasal area is highly needed in the production of some *Yoruba* sounds, this calls for a proper studying of the nasal cavity. In the production of sounds, the velum is the opening to the nasal tract of the nasal cavity, so if the velum is closed up, no air flow to the nasal cavity. The mass of pressure released by the palate of the nasal cavity is assumed to have a negligible mass and weight in consideration to the passage of air flow through the entire nasal cavity for the release of air through it. A nasal vowel is a vowel that is produced with a lowering of the velum so that air passes both through the nose as well as the mouth. In *Yoruba* language, the nasal vowels are *en ,in ,on ,un*.

2.1 System Models

A system is a mathematical abstraction that is devised to serve as a model for the phenomenon, in terms of mathematical relations, that exists among the input, the output and the state variables. Modeling is the art of formulating application in terms of precisely described, well-understood problems. Proper modeling is the key to applying algorithmic design techniques to real-world problems.

Models challenge the robustness of prevailing theory through perturbations and they also reveals the apparently simple to be complex and complex to be simple [18]. Mathematical models can be static or dynamic, explicit or implicit, linear or non-linear, discrete or continuous, deterministic or probabilistic.

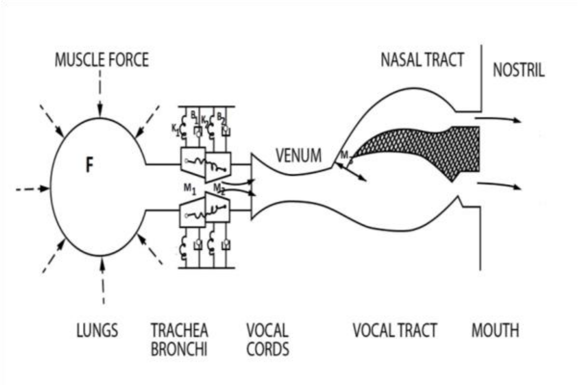


FIGURE 6: Physical and Mechanical Representation of Model Source: [22].

2.2 Mathematical Model

Dynamism, uncertainty and complexity is dominate in today's world. Dynamics are central to understanding complex and interacting systems. Many researchers developed their mathematical models from the mechanical model. The mechanical model always depicts and give a simple, yet efficient picture and analysis of voice sounds through the vocal chords. Mathematical models are usually composed of variables having relationships which are in the form of operators which depicts the behaviour of real objects. Mathematical models are used because it is very precise language and this helps to formulate ideas and identify the underlying assumptions, it is also a concise language with well-defined rules for manipulations and since computers can be used to perform numerical computations.

According to [36], mathematical models of vocal folds serves as a viable alternative to direct experiment studies using strobovideolaryngoscopy or electroglottography techniques or methods. In speech production, the vocal tract, which is the major speech organ that allows the air to flow in series of pulses, is where the masses were been located, with each noted point moving in a springlike nature with an attached damper. Researchers started with a single mass mechanical model and then moved to two-mass mechanical model. The two-mass model shows one vocal fold by two coupled oscillators. Each of the oscillator consisting of a spring stiffness, a mass and a damper. Mass M_1, linear spring stiffness K_1 and B_1 represent the damping coefficient of the first part of the vocal fold. Mass M_2, linear spring stiffness K_2 and B_2 represent the the damping coefficient of the second part of the vocal fold. The displacement-dependent driving forces F_1 and F_2 are proportional to the average acoustic pressures in the two sections of the glottis. Whenever a section is closed (due to the collision of its sides) the corresponding driving force is zero. Note that it is these forces that provide the feedback of the acoustic pressures to the mechanical system.

The dynamic response of the mechanical part of the two-mass model can be described by the equations of motion of the two mass is given as follows:

$$M_1\ddot{x}_1(t) + B_1\dot{x}_1(t) + K_1x_1(t) + K_c(x_1 - x_2) = F_1(x_1) \quad (4)$$

$$M_2\ddot{x}_2(t) + B_2\dot{x}_2(t) + K_2x_2(t) + K_c(x_2 - x_1) = F_2(x_2) \quad (5)$$

where x_1 is the displacement of the first mass, x_2 is the displacement of the second mass, the stacked dot $\dot{x}_1$ and $\dot{x}_2$ denote the first order derivative of x_1 and x_2 with respect to the temporal variable t respectively, also the stacked dots $\ddot{x}_1$ and $\ddot{x}_2$ denote the second order derivative of x_1 and x_2 with respect to the temporal variable t respectively.

Equation (3.1) and (3.2) can be rephrased as

$$M_1\ddot{x}_1(t) + B_1\dot{x}_1(t) + (K_c + K_1)x_1(t) - K_cx_2(t) = F_1(x_1) \quad (6)$$

$$M_2\ddot{x}_2(t) + B_2\dot{x}_2(t) + (K_c + K_2)x_2(t) - K_cx_2(t) = F_2(x_2) \quad (7)$$

From the above equations, it is observed that the formula followed a particular pattern in which is the mass is the second-order differential of the displacement, the damping coefficient having a relationship function with the first order differential and the spring constant relating with the displacement. More masses can be added to refine the synthesis of flow in the vocal fold, and which formed three-mass, four-mass, five-mass, and then M-mass models. In formulating a M-mass model, one must put into consideration that K_c is caused by the lateral movement of the interaction of the adjacent masses. Therefore, for a two-mass model, it is $K_c = K_{1,2}$ and that is a relationship that extends through the masses that are adjacent to each other. Similarly, one can formulate the system of equations of the dynamics of M-mass model as

$$M_1\ddot{x}_1 + B_1\dot{x}_1 + K_1x_1 + K_{1,2}(x_1 - x_2) = F_1 \quad (8)$$

$$M_i\ddot{x}_i + B_i\dot{x}_i + K_{i,i-1}(x_i - x_{i+1}) + K_{i,i+1}(x_i - x_{i+1}) = F_1, i = 2,...,M - 1 \quad (9)$$

$$M_M\ddot{x}_M + B_Mx_M\dot{} + K_Mx_M + K_{M,M-1}(x_M - x_{M-1}) = F_1 \quad (10)$$

where K_i denotes the spring stiffness of mass M_i, B_i represent the damping coefficient of mass M_i and $K_{i,j}$ represent the coupling spring stiffness between mass M_i and mass M_j, where j=i+1 or j=i-1 depending on the positioning.

The equations of motion of the M-mass model can be written in matrix form as

$$M_{mass}{\sim}\ddot{x} + B_{mass}{\sim}\dot{x} + K_{mass}{\sim}x = F_{mass}{\sim} \quad (11)$$

According to [17], the addition of more masses beyond two-mass at the oral cavity of the vocal chord does not necessarily fully describe the flow inside the vocal fold. Therefore, one can use the two-mass to efficiently represent the oral cavity of the speech production organ. There is need to include the nasal cavity model, and to formulate the mechanical mode of the nasal cavity, a mass is placed at the entrance to the nasal cavity, that is the velum. The mass of the oscillator in the nasal cavity is a representation of the mass of the velum. The nasal cavity is represented as

$$M_N\ddot{x}_N + B_N\dot{x}_N + K_N x_N + K_c(x_{1,2..,M,N}) = F_N x_N \tag{12}$$

The vibration of vocal chords produces the primary source of sounds for vowels [29]. The first and the second masses are positioned at the vocal chord while the third is positioned at the opening of the nasal cavity(which is the velum). The vocal chord can be studied as an isolated system since the configuration of the vocal tract has no effects on its dynamics [19]. The resulting mathematical model is

$$M_1\ddot{x}_1(t) + B_1\dot{x}_1(t) + K_1x_1 + K_{12}(x_1(t) - x_2(t)) = F_1(x_1) \tag{13}$$

$$M_2\ddot{x}_2(t) + B_2\dot{x}_2(t) + K_2x_2(t) + K_{12}(x_2(t) - x_1(t)) = F_2(x_2) \tag{14}$$

$$M_3\ddot{x}_3(t) + B_3\dot{x}_3(t) + K_3x_3 + K_{123}(x_1,x_2,x_3)(t) = F_3(x_3) \tag{15}$$

where $F_3(0)$ will be zero when a sound is orally produced without nasality. The parameters in the model are M_1, M_2, M_3, B_1, B_2, K_1, K_2, F_1, F_2, B_3, K_{12}, K_{123}, F_3. These parameters and the corresponding way they relate to speech organs are shown in Table 3.1. K_{123} is a non-linear parameter that relate the spring constant between the two masses at the oral cavity and the spring constant of the velum at the nasal cavity.

$$\begin{pmatrix} M_1\dfrac{d^2}{dt^2} & B_1\dfrac{d}{dt} & K_1 \\ M_2\dfrac{d^2}{dt^2} & B_2\dfrac{d}{dt} & K_2 \\ \vdots & \vdots & \vdots \\ M_m\dfrac{d^2}{dt^2} & B_m\dfrac{d}{dt} & K_m \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_m \end{pmatrix} + \begin{pmatrix} K_c^1 \\ K_c^2 \\ \vdots \\ K_c^m \end{pmatrix} = \begin{pmatrix} F_1 \\ F_2 \\ \vdots \\ F_m \end{pmatrix}$$

These equation become problematic if $M > 3$,

$$\text{1st: } \alpha X_1 + \beta X_2$$

$$\text{2nd: } \alpha_i X_{i-1} + \beta_i X_i + \gamma_i X_{i+1}$$

$$\forall\ i=2,...,m-1$$

The mathematical model to be implemented is shown in the equation below which is in matrix form:

$$\begin{pmatrix} \alpha^1 + M_1 & \beta\ {}^1 + B_1 & K_1 \\ \alpha^2 + M_2 & \beta\ {}^2 + B_2 & \gamma^2 + K_2 \\ M_3 & \beta\ {}^3 + B_3 & \gamma^3 + K_3 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} = \begin{pmatrix} F_1 \\ F_2 \\ F_3 \end{pmatrix}$$

M_3 is assumed to be zero, since the mass of velum is very insignificant in relation to the other masses M_1 and M_2. K_{123} which represent the non-linear spring constant interactions between the two masses in the oral cavity (that is, at the vocal chord) and the mass at the velum, is one of the main significant part of these model which does not exists in the oral cavity model that has been modeled in existing literatures.

2.3 Electrical Model

The electrical model was formulated from the mechanical models, which is the representation of the speech production organs. There is need to know the corresponding equivalent signals, symbols and quantities of these mathematical and mechanical representation in electrical forms for the acoustic model to be designed. The electrical/acoustic model gotten from the mathematical model formulated for nasal voiced speech is shown in Figure 7. This work focused on the mathematical model and not the electrical model representations.

2.4 Design of Computational Model

A computational model is a mathematical model in computational science that requires the use of extensive computational resources to study the behavioural pattern of a complex system by computer simulation. Computational model makes use of variables and parameters which are governed by mathematical relationships that characterised the system been understudied. Computational models are created to simulate a set of processes observed in the natural world in order to gain an understanding of the processes and to predict Table 1: Parameters used in the

S/N	Variable	Variable Description	Units
1.	M1	Mass at the first point in the vocal chord	g
2.	M2	Mass at the second point in the vocal chord	g
3.	M3	Mass of the velum at the nasal cavity	g
4.	X1	Displacement of masss at the first point of the vocal chord	m
5.	X2	Displacement of masss at the first point of the vocal chord	m
6.	X3	Displacement of masss at nasal cavity	m
7.	K12	Coupling spring stiffness between the first and the second mass	N/m
8.	K123	Coupling spring stiffness between the oral and nasal cavity	N/m
9.	K1	Linear spring constant of the first mass	N/m
10.	K2	Linear spring constant of the second mass	N/m
11.	K3	Linear spring constant of the mass at nasal cavity	N/m
12.	B1	Damping coefficient for the first mass	Ns/m
13.	B2	Damping coefficient for the second mass	Ns/m
14.	B3	Damping coefficient for the mass at nasal cavity	Ns/m
15.	F1	Force at the first mass	Ns/m
16.	F2	Force at the second mass	Ns/m
17.	F3	Force at the nasal cavity	Ns/m

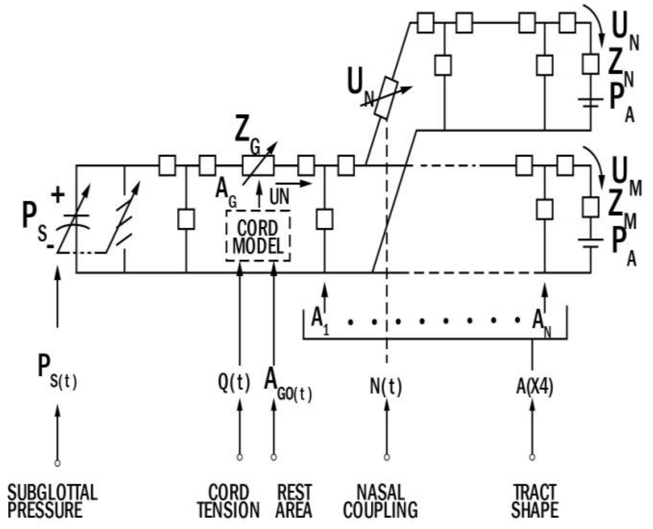


FIGURE 7: Electrical Circuit of Model Source: [21].

model are the outcome of natural processes given a specific set of input parameters. The work started from mechanical model, because the best way the movement of speech articulatory organs can be depicted is through the mechanical mechanism.

The Algorithm for designing the model is:

Step 1: Start

Step 2: Input the values of the input parameters M1, M2, M3, K1, K2, K3, K12, K123, F1, F2, F3, B1, B2, B3

Step 3: Let y(1) represent x1, y(2) represent x2, y(3) represent x3, y(4) represent $\frac{dx_1}{dt}$, y(5) represent $\frac{dx_2}{dt}$, y(6) represent $\frac{dx_3}{dt}$

Step 4: Set the initial value of the displacements at the three points to 0 (since the masses have zero displacement before any speech was produced)

Step 5: Solve using numerical solution

Step 6: Call Ode45s function solver

Step 7: Plot time versus x1 and time versus x3

Step 8: End

2.5 Solving Ordinary Differential Equation

The need of technical computing since around 1980 necessitated the urgent and high need for matlab. MATLAB has a large library of tools that can be used to solve differential equations. Solving a system of ODE in MATLAB is quite similar to solving a single equation, though the system of equations can not be defined as an inline functions we must define it as an M-file. MATLAB has a staggering array of Table 2: Ordinary Differential Equation Solvers

S/N	Solver	Problem Type	Order of Accuracy	When to use
1.	ode45	Nonstiff	Medium	Most of the time. This should be the first solver you try. Basedon explicit Runge-kutta method.
2.	ode23	Nonstiff	Low	For problems with crude error tolerances or for solving moderately stiff problems.
3.	ode113	Nonstiff	Low to high	For problems with stringent error tolerances or for solving computationally intensive problems. Multistep solver.
4.	ode15s	Stiff	Low to medium	if ode45 is slow because the problem is stiff. Uses a variable order method.
5.	ode23s	Stiff	Low	if using crude error tolerances to solve stiff systems and the mass matrix is constant. One step solver.
6.	ode23t	Stiff Moderately	Low	For moderately stiff problems if you need a solution without numerical damping.
7.	ode23tb	Stiff	Low	If using crude error tolerances to solve stiff systems. Often more efficient than ode15s.

tools for numerically solving ordinary differential equations. The table 2 shows the lists of several solvers and their properties. Some Ordinary Differential Equations (ODE’s) are referred to as ”stiff” in that the equation includes terms that can lead to rapid variation in the solution and thus produce instabilities in using numerical methods because the errors compound dramatically over time. For most ”nonproblematic” ODEs, the solver ode45 works quite well and should be the initial choice. The numerical methods will be able to handle both linear and nonlinear equations. Therefore, for the implementation of this model, ode45 was used.

3 RESULTS AND DISCUSSIONS

The mathematical model that was developed was solved using ordinary differential equation. As many literatures were studied, those literatures that has articulatory parameters used in there studies only have the oral parameters. [28] was able to do a critical analysis of the articulators in production of speech. The Table 3: Parameters for Articulatory Organs

S/N	Parameter Label	Category 1	Category 2	Category 3	Category 4	Category 5
1.	kl	80.000	100.0000	80.000	80.000	112.250
2.	k2	8.000	100.0000	8.000	8.000	157.140
3.	k3	8.000	8.0000	8.000	8.000	8.000
4.	y01	0.000	0.0000	0.000	0.000	0.000
5.	y02	0.000	0.0000	0.000	0.000	0.000
6.	y03	0.000	0.0000	0.000	0.000	0.000
7.	k12	25.000	31.2500	25.000	25.000	6.450
8.	k123	25.000	25.0000	25.000	25.000	25.000
9.	y012	0.000	0.0000	0.000	0.000	0.000
10.	b1	0.010	0.1000	0.010	0.020	0.100
11.	b2	0.600	0.6000	0.010	0.020	0.600
12.	b3	0.600	0.6000	0.600	0.600	0.600
13.	F1	8000.000	8000.0000	8000.000	8000.000	8000.000
14.	F2	8000.000	8000.0000	8000.000	8000.000	8000.000
15.	F3	8000.000	8000.0000	8000.000	8000.000	8000.000
16.	m1	0.1250	0.1563	0.1250	0.125	0.058
17.	m2	0.025	0.0313	0.025	0.025	0.082
18.	m3	0.125	0.1250	0.125	0.125	0.125

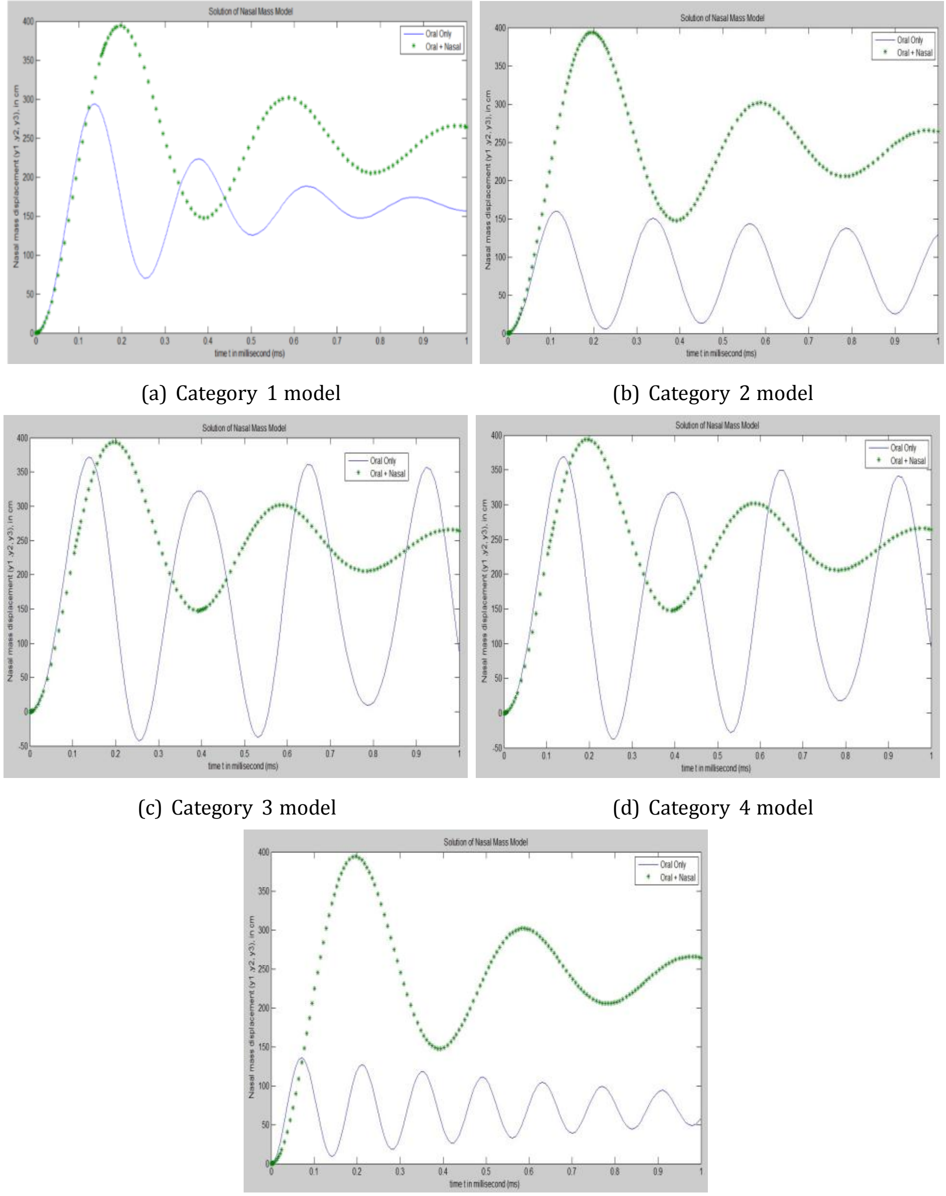
articulatory parameter of the nasal cavity was formulated by knowing the mass of velum, the damping capacity of velum and its vibrating prowess. So, these nasal cavity parameters were combined with the variously used oral cavities to form the parameters we used. The parameters in all of the understudied literatures with parameters were categorized into five categories in relation with the closeness of values used as parameters.

The first category which is the parameters from [9, 4, 16, 17] all agreed in their values for oral cavity. These nasal cavity parameter was combined with all the known oral cavity parameters, since the oral cavities has been worked upon before, so data can be gotten from there and the result was tested and shown. The combination of data was done because the data has to be gotten from various literatures. The parameters of Category 1 shown in Table 3 were modified by inputting the nasal cavity parameters into that of [9, 4, 16, 17] and then simulated with the result shown in Figure 8 (a).

The second category is a set of parameters which were used in [7, 8] with the modified parameters formulated shown in Category 2 column of Table 3 and the computational model shown in Figure 8 (b).

The third category used parameters in [39] the column Category 3 as shown in Table 3 and while the fourth category used in [26] is column Category 4, as shown in Table 3 are similar except for the damping coefficients of the vocal chord at the two points considered which have different values. For [39], the damping coefficients are 0.01 while that of [26], the damping coefficients are 0.02. With these difference in the damping coefficients the simulated results did not show any remarkable change as shown in Figure 8(c) and Figure 8(d). Also, as could be seen in the two figures, the damping coefficient value between the oral cavity parameters and the parameter at the nasal cavity been at different significant level made the oscillations at the oral moving at a great damping level.

The fifth and last category in [29] also used some set of parameters which are orals and the corresponding nasal parameters formulated in column Category 5, as shown in Table 3 and the computational model as



(a) Category 1 model

(b) Category 2 model

(c) Category 3 model

(d) Category 4 model

(e) Category 5 model

FIGURE 8: Nasal mass displacement-time plot.

TABLE 4: Simulation Results.

Categories	Oral Frequency (KHz)	Nasal Frequency (KHz)	Area under cavity Nasal graph(cm^3ms)	Area under Oral cavity graph(cm^3ms)	Volume of sound in oral cavity(cm^3)	Volume of sound in nasal cavity(cm^3)
Category 1	4.00	2.50	161.3651	238.1008	645.460400	403.412750
Category 2	4.50	2.50	079.1639	238.1028	356.237550	595.257000
Category 3	4.00	2.50	170.7075	238.1059	682.830000	595.264750
Category 4	4.00	2.50	170.0592	238.1060	680.236800	595.265000
Category 5	7.05	2.55	069.7595	238.1025	491.804475	607.161375

seen in Figure 8(e). In [29], the first and second mass has a low value and with the high value of spring constant at both masses, made the oscillations at the oral level so low. So, the production of oral voice speech is highly dependent on the mass of the vocal chord.

An individual with weak vocal cord cannot produced a clear oral sounds, even though nasal sounds can be well spoken by the individual. Whenever damping coefficient is constancy over the points positioned with masses, even if vocal chords of individuals have different values, yet the speech production is still the same and normal.

The frequency of each of the graph for each of the plot, that is, for oral and nasal cavities plots were gotten, since frequency is known to be the number of cycles that is completed in a certain amount of time.

$$\text{Frequency} = \frac{\text{Number of cycles}}{\text{Amount of time}} \qquad (16)$$

Using equation 4.1, the frequency obtained is shown as seen in Table 4. So also the area under the graph can be calculated which gives volume-second of the model. The area under the graph can be obtained from the matlab graph through a line of code.

To get the volume of air in the production of the sounds, multiply the frequency of sounds with the volume-second (that is the area under the graph) and for the five categories considered the results is shown in Table 4.

$$\text{Area under graph } (cm^3\text{s}) = \text{Volume } (cm^3) * \text{Time(s)} \qquad (17)$$

$$\text{Volume of sound } (cm^3) = \text{Area under graph } (cm^3\text{s}) * \text{Frequency of sound waves(Hz)} \qquad (18)$$

The response time for all the five category were done and studied as shown in figure 5. It was found that the rise time, settling time, settling maximum, settling minimum, overshoot, undershoot, peak and peak time for category 1,3 and models are all the same. Though the damping coefficient has different value over these three categories, yet the response rate stayed constant. It was discovered that the spring constant determines the response rate, as the spring constant changes, the response rate also changes.

4 SUMMARY AND CONCLUSION

When it comes to a language like *Yoruba*, nasality and tonality are very important in the articulation of the speech. These gives the correct meaning to any sound produced in the *Yoruba* language. In standard *Yoruba* language, there are seven vowels and five nasal vowels in Standard *Yoruba* language, they are: *a, e, e., i, o, o., u* and *an, e.n, in, o.n, un*.

TABLE 5: Response Time.

Categories	Rise Time	Settling Time	Settling Minimum	Settling Maximum	Overshoot	Under shoot	Peak	Peak Time
Category 1	0.9000	5.9800	6.4000e+04	3.2000e+05	400.0000	0	3.2000e+05	5
Category 2	0.9845	5.9800	6.4000e+04	2.5559e+05	299.3610	0	2.5559e+05	5
Category 3	0.9000	5.9800	6.4000e+04	3.2000e+05	400.0000	0	3.2000e+05	5
Category 4	0.9000	5.9800	6.4000e+04	3.2000e+05	400.0000	0	3.2000e+05	5
Category 5	0.4176	5.9559	6.4000e+04	1.3793e+05	115.5172	0	1.3793e+05	4

This study developed a model for the physical and human production of voiced sounds which is translated into its mechanical model approach of human speech production system representation. The spring-mass damper system was able to model the vocal chord articulation in voice production. The corresponding acoustic and electrical circuit was been developed which was later used to formulate the mathematical model of speech sounds. This research therefore presents a computational model for articulatory mechanism of *Yoruba* speech voiced.

Although the human voice production system is complex and dynamic in its nature. Considering the fact that the positions of the articulators changes per time. The physical, electrical and mathematical model of the articulatory dynamics of voiced sound speech had been studied and it was found that all the existing models did not put into consideration the nasality of sound speech. These particular research was then able to formulate the computational model involving the nasal cavity and the formulated model was designed. The implementation of the design was done on Matlab. We have been able to establish the computational model for articulatory mechanism of *Yoruba* voiced speech as presented in this research. The results have established that more volume of velocity of air will be needed for nasal vowels than oral vowels in the production of *Yoruba* voiced speech. We are able to establish a computational model for nasalized *Yoruba* vowels. The results of these research provides a better understanding of the human speech articulatory system and mechanism in the production of *Yoruba* voiced speech and the model developed serve as resources for *Yoruba* speech recognition and text-to-speech applications. The modeling of articulatory mechanism for unvoiced speech, that is the consonants sounds, and syllables are the area of further research work in which the principle in this work could be extended.

5 REFERENCES

[1] A. Aalto, D. Aalto, J. Malinen, and M. Vainio. Interaction of vocal fold and vocal tract oscillations. In *24th Nordic Seminar on Computational Mechanics, J. Freund and R. Kouhia (Eds.), Aalto University*, pp. 1 – 4, 2011.

[2] L. A. Akanbi, and O. A. Odejobi. Automatic recognition of oral vowels in tone language: Experiments with fuzzy logic and neural network models. Applied Soft Computing, 11:1467 – 1480, 2011.

[3] A. Akinlabi. Understanding Yorubalife and culture: The sound system of Yoru`b´a. Africa World Press Inc., Eritrea, 2004.

[4] D. A. Berry. Mechanisms of modal and nonmodal phonation. Journal of phonetics, 29:431 – 450, 2001. [5] A. Breen. Speech synthesis model: A review. Electronics and Communications Engineering Journal, February, 1992:19 – 31, 1992.

[5] R. Bronson, and G. Coster. Differential equations. Schaum’s outline series McGraw, United States, ISBN: 978-0-07-161162-6, 3rd edition, 2006.

[6] E. Cataldo, F. R. Leta, J. Lucero, and L. Nicolato Synthesis of voiced sounds using low-dimensional models of the vocal cords and time-varying subglottal pressure. Mechanics Research Communications, 33(6):250 – 260, 2006 a.

[7] E. Cataldo, J. C. Lucero, R. Sampaio, and L. Nicolato. Comparison of Some Mechanical Models of Larynx in the Synthesis of Voiced Sounds. Journal of the Brazil Society of Mechanical Science and Engineering, XXVIII(4):461–466, 2006 b.

[8] E. Cataldo, C. Soize, C. Desceliers, and R. Sampaio. Uncertainties in mechanical models of larynx and vocal tract for voice production. In XII International Symposium on Dynamics Problems of Mechanics, Brazil, pp. 1 – 10, 2007.

[9] K. E. Cummings, and M. A. Clements. Glottal models for digital speech processing: A historical survey and new results. Digital signal processing, 5:21 – 42, 1995.

[10] L. Cveticanin. Review on Mathematical and Mechanical Models of the Vocal Cord. Journal of Applied Mathematics, 2012:1–18, 2012.

[11] H. K. Dass. Advanced engineering mathematics. 81-219-0345-9. S. Chand & company ltd., New Delhi, India. ISBN: 81-219-0345-9, 17th edition, 1988.

[12] P. Dikshit. ”An algorithm for locating fundamental frequency (f0) markers in speech.”, Master’s thesis, Department of Computer Engineering, Old Dominion University, United States, 2004.

[13] G. Fant. Speech production: A voice source dynamics. STL-QPSR, 2 - 3:17 – 37, 1980.

[14] J. Flanagan, and L. Landgraf. Self-oscillating source for vocal-tract synthesizers. IEEE Trans. On Audio and Electroacoustics, 16:57 – 64, 1968.

[15] J. Huang. Articulatory speech synthesis and speech production modelling. Doctor of philosphy thesis, University of Illinois, Urbana-Champaign, United States, 2001.

[16] K. Ishizaka, and J. L. Flanagan. Synthesis of Voiced Sounds from a two-mass model of Vocal Cords. The Bell system Technical Journal, 51(6):1233 – 1268, 1972.

[17] M. E. Joshua. Why Model? In Sfi Working Paper, pp. 1–6, 2008.

[18] J. C. Lucero, K. G. Lourenco, N. Hermant, A. V. Hirtum, and X. Pelorson. Effect of source-tract acoustical coupling on the oscillation onset of the vocal folds. Journal Acoustic Society of America, 132(1):403 – 411, 2012.

[19] O. A. Od´e.jo.b´ı. Articulatory organs in speech production. Personal collections retrieved February,2015a.

[20] O. A. Od´e.jo.b´ı. Electrical Circuit of Model. Personal collections retrieved February,2015b.

[21] O. A. Od´e.jo.b´ı. Physical and Mechanical Representation of Model. Personal collections retrieved February,2015c.

[22] T. O. Olorunfemi. Development of a computational model for predicting acoustic data from articulatory configuration of standard Yorubavowels. Master’s thesis, Department of Computer Science and Engineering, Obafemi Awolowo University, If´e., Nigeria, 2013.

[23] P. Palo. A review of articulatory speech synthesis. Master’s thesis, Helsinki University of Technology, Finland, 2006.

[24] E. Rank. Oscillator-plus-noise modelling of speech signals. Doctor of enginering’s thesis, Vienna University of Technology, Austria, 2005.

[25] M. F. Regner, C. Tao, D. Ying, A. Olszewski, Y. Zhang, and J. J Jiang. The Effect of Vocal Fold Adduction on the Acoustic Quality of Phonation: Ex Vivo Investigations. Journal of Voice, 26(6):698 – 705, 2012.

[26] N. Ruty, A. V. Hirtum, X. Pelorson, I. Lopez, and A. Hirschberg. A mechanical experimental setup to simulate vocal folds vibrations. ZAS Papers in Linguistics, 40:161–175, 2005.

[27] C. Scully. Linguistic units and units of speech production. Speech Communication, 6:77 – 142, 1987.

[28] B. H. Story. An overview of the physiology, physics and modeling of the sound source for vowels. Acoustic Science & Technology, 23(4):195 – 206, 2002.

[29] B. H. Story, and I. R. Titze. Voice simulation with a body-cover model of the vocal folds. Journal Acoustic Society of America, 97(2):1249 – 1260, 1995.

[30] K. A. Stroud, and D. J. Booth. Adanced engineering mathematics. 1-4039-0312-3. Palgrave Macmillian Publisher, London. ISBN: 1-4039-0312-3, 4th edition, 1995.

[31] K. A. Stroud, and D. J. Booth. Engineering mathematics. 0-333-91939-4. Palgrave Macmillian Publisher, London. ISBN: 0-333-91939-4, 5th edition, 2001.

[32] A. J. S. Teixeira, R. Martinez, L. N. Silva, L. M. T. Jesus, J. C. Principe, and F. A. C. Vaz (2005). Simulation of human speech production applied to the study and synthesis of European Portuguese. EURASIP Journal on Applied Signal Processing, 9:1435 – 1448, 2005.

[33] I. R. Titze. On the mechanics of vocal-fold vibration. Journal Acoustic Society of America, 60(6):66 – 80, 1977.

[34] I. R. Titze, and B. H. Story. Rules for controlling low-dimensional vocal fold models with muscle activation. Journal Acoustic Society of America, 112(3):1064 – 1076, 2002.

[35] J. Xin, and Y. Qi. Mathematical modeling and signal processing in speech and hearing sciences, volume 10. Springer international publisher, Switzerland, 2014.

[36] H. Yehia, and F. Itakura. A method to combine acoustic and morphological constraints in the speech production inverse problem. Speech Communication, 18:151 – 174, 1996.

[37] H. Yehia, P. Rubin, and E. Vatikiotis-Bateson. Quantitative association of vocal-tract and facial behavior. Speech Communication, 26:23 – 43, 1998.

[38] Y. Zhang, J. J. Jiang, L. Biazzo, and M. Jorgensen. Perturbation and Nonlinear Dynamic Analyses of Voices from Patients with Unilateral Laryngeal Paralysis. Journal of Voice, 19(4):519 – 528, 2005.

[39] M. K. Reddy, and K. S. Rao. Excitation modelling using epoch faetures for statistical parametric speech synthesis. Computer Speech and Language, 60(2020):10 – 29, 2020.

[40] S. P. Panda, A. K. Nayak, and S. C. Rai. Survey on speech synthesis techniques in Indian languages. Multimedia Systems, 26:453 – 478, 2020.

[41] L. Juvela, B. Bollepalli, V. Tsiaras and P. Alku. Glotnet - a raw waveform model for the glottal excitation in statistical parametric speech synthesis. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 27(6):1019 – 1030, 2019.

[42] O. Perrotin, and I. McLoughlin. A spectral glottal flow model for source-filter separation of speech. International Conference on Acoustics, Speech and Signal Processing, 2019:7160 – 7164, 2019.

[43] T. Kenter, V. Wan, C. Chan, R. Clark and J. Vit. CHiVE: Varying prosody in speech synthesis with linguistically driven dynamci hierarchical conditional variational network. International Conference on Machine Learning, 2019:3331 – 3340, 2019.

[44] A. Indumathi and E. Chandra. Survey on Speech Synthesis. Signal Processing: An International Journal,6(5): 140 – 145, 2012.

[45] Z. Mnasri, F. Boukadida, and N. Ellouze. F0 Contour Modeling for Arabic Text-to-Speech Synthesis usinf Fujisaki Parameters and Neural Networks. Signal Processing: An International Journal,4(6): 352 – 369, 2011.

[46] M. B. Chandak, R. V. Dharaskar and V. M. Thakre. Text to Speech Synthesis with Prosody Feature: Implementation of emotion in speech output using forward parsing. International Journal of *Computer Science and Security*,4(3): 352 – 360, 2010.