Modeling laryngeal muscle activation noise for low-order physiological based speech synthesis

Rodrigo Manríquez¹, Sean D. Peterson², Pavel Prado¹, Patricio Orio³, Matías Zañartu¹

¹Universidad Técnica Federico Santa María, Chile
²University of Waterloo, Canada
³Universidad de Valparaíso, Chile

rodrigo.manriquezp@alumnos.usm.cl, peterson@uwaterloo.ca, pavel.prado@usm.cl, patricio.orio@uv.cl, matias.zanartu@usm.cl

Abstract

Physiological-based synthesis using low order lumped-mass models of phonation have been shown to mimic and predict complex physical phenomena observed in normal and pathological speech production, and have received significant attention due to their ability to efficiently perform comprehensive parametric investigations that are cost prohibitive with more advanced computational tools. Even though these numerical models have been shown to be useful research and clinical tools, several physiological aspects of them remain to be explored. One of the key components that has been neglected is the natural fluctuation of the laryngeal muscle activity that affects the configuration of the model parameters. In this study, a physiologically-based laryngeal muscle activation model that accounts for random fluctuations is proposed. The method is expected to improve the ability to model muscle related pathologies, such as muscle tension dysphonia and Parkinson’s disease. The mathematical framework and underlying assumptions are described, and the effects of the added random muscle activity is tested in a well-known body-cover model of the vocal folds with acoustic propagation and interaction. Initial simulations illustrate that the random fluctuations in the muscle activity impact the resulting kinematics to varying degrees depending on the laryngeal configuration.

Index Terms: Speech synthesis, vocal folds, voice, muscle activation

1. Introduction

Lumped-element models of voiced speech are capable of investigating a wide range of controlled scenarios at minimal computational cost. These models can mimic and predict complex physical phenomena observed in speech production and have been shown to be useful tools for the investigation, diagnosis, and treatment of voice disorders [1][2]; furthermore, they have received significant attention due to their ability to efficiently perform comprehensive parametric investigations that are cost prohibitive with more advanced computational tools [3]. Recently, these reduced order models have been recently shown advantageous in that they can yield data that are difficult or impossible to clinically acquire, by means of Bayesian estimation methods [4]. In spite of these notable advancements, several physiological aspects of these reduced order models remain to be incorporated, particularly when describing their connection with the nervous system.

Reduced order vocal fold (VF) models are constituted by a set of coupled components, namely a lumped element description of the VFs, an analytical solution of the glottal flow behavior, and a plane wave representation of the sub and supra glottal acoustic fields. Herein, the VF model is typically configured using physiological rules of muscle activation [5], that allow for a meaningful construction of the VF model parameters. However, the activation rules are based upon several untested assumptions and neglect any neural description of the laryngeal muscle activity. The lack of such neural descriptions results in perfectly constant muscle activations and unnatural sound quality for the synthetic voices.

In this study, a physiologically-based laryngeal muscle activation scheme that accounts for random fluctuations is presented. The proposed method is expected to improve both the physiological relevance of the overall speech production model and the ability to model muscle related pathologies, such as muscle tension dysphonia and Parkinson’s disease. At the same time, natural random fluctuations in the model parameters are expected to enhance the resulting sound quality of the voice synthesizer.

2. Methods

2.1. Physiological and morphological aspects of muscular activation

From a neural perspective, the basic unit of a muscle is the motor unit (MU), which consists on a motor neuron and muscle fiber innervated by the axon of the neuron. Large muscles with wide and gross movements have thousands of fibers per motor neuron, while muscles that perform precise contractions have fewer fibers per motor neuron. When a neuronal signal is sent, a MU action potential is generated, which stimulates the fibers to contract synchronously. The electric impulse generated, which corresponds to the sum of all action potentials, is known as motor-unit action potential.

For each spike that stimulates a fiber, a single twitch (or fiber contraction) is produced on them, lasting for a fraction of time. Successive twitches may add up to generate a stronger action [6]. Figure 1 illustrates the temporal summation effect of successive twitches. If spikes are fired at a higher rate, the resulting contraction is larger, thus having a wave summation effect, resulting in temporal variability. This effect is also illustrated in Figure 1, in which neuronal variability can be observed at a single fiber level.

Amplitude and temporal properties of a twitch depend on the type of fiber that comprises a MU. In our scheme, all muscle fibers in a MU are assumed to be of the same type, and although there are many types of fibers, they can be simplified into two principal groups: Type I or slow, and Type II or fast [7]. Slow fibers are fatigue-resistant, with the smallest force...
or twitch tension and slowest contraction. Fast fibers can also be fatigue-resistant, with large forces and faster contraction, or easily fatigable but having the largest force and fastest contraction. Each GMU (labeled by \( j = 1, ..., N_a \)) is firing at a different rate \( F_j \), as noted by Märtensson [12]. A single GMU can only fire at a maximum rate of \( F_{\text{max}} \). If \( F_j > F_{\text{max}} \), then \( F_j \) is set at \( F_{\text{max}} \). Logically, this is applied after the recruitment step, so there may be some GMUs firing at \( F_{\text{max}} \) while there are others below this value.

![Figure 2: Fast (right) and slow (left) Twitch waveforms. Each time a MU fires a spike, a corresponding twitch is triggered as a response. The sum of successive twitches is the basis of the wave summation model.](image)

In an effort to capture the inherent variability in neural firing, and thus the contraction response, a degree of randomness is also added to various components. First, a rule of five, each time a new GMU is recruited, the firing rate for the new GMU is set at a value between 3 to 7 (normal distribution with a mean of 5 and a standard deviation of 2), instead of 5 Hz fixed. Also, it is known that intervals between spikes are not regular. Based upon the results of Moritz et al [13] that describes the inter spike interval (ISI) in muscular activation, a normally distributed ISI with a coefficient of variation (CV) of 0.2 was considered.

We now normalize the resulting force of contraction to describe the resulting muscle activation. Thus, the activation \( a_m \) for a given muscle \( m \) is defined as follows:

\[
a_m(F, N) = \frac{p_m(F, N)}{p_{\text{max}}(F_{\text{tet}}, N)} \tag{5}
\]

in which \( p_m \) is the time series for the force of contraction with random components (note the bold notation to differentiate from \( p_{\text{m}} \), which has no random components). For the normalization step, the model described in equation 1 is used (without any random components). The muscle activation \( a_m \) is normalized, so the muscle is considered fully activated (or tetanized) if \( a_m = 1 \). Conversely, a muscle has no activation related if \( a_m = 0 \), which has no firing rate associated. However, this situation is never used because a fully relaxed muscle has a residual muscle tension. \( F_{\text{tet}} \) is the fire rate at which all GMUs are firing at maximum capacity, considering \( p_m \) (with no random components). \( F_{\text{tet}} \) should be higher that \( F_{\text{max}} \) due to the rule of five.

### 2.3. Model Parameters

The model proposed can be used to simulate any muscle in which muscle activation is required. In this case, two muscles were simulated: the thyroarytenoid (TA) and the lateral cricoarytenoid (LCA), due to their importance in voice phonation. Both TA and LCA are considered fast muscles, with 10% of slow fiber and 90% of fast fiber [12]. The parameters related to the model are presented in Table 1.

### 2.2. Wave summation model

MUs are recruited into group of motor units (GMUs), where we assume that all GMUs are identical for a given muscle: that is, they have the same proportion of slow and fast MUs. In addition, the rule of five is assumed for the recruitment of GMUs, in which each subsequent GMU is recruited if an active GMU increases its firing rate by approximately 5 [Hz] [11]. Considering \( N \) Groups of MUs, in which the first one is firing at \( F \) [Hz], then the set of equations that governs the recruitment of GMUs are:

\[
p_m(F, N) = \sum_{j=1}^{N_a} \sum_{k=0}^{\infty} n_s \cdot p_s(t - k/F_j) + \sum_{j=1}^{N_a} \sum_{k=0}^{\infty} n_f \cdot p_f(t - k/F_j) \tag{1}
\]

\[
p_{s,f}(t) = \frac{t}{\tau_{s,f}} e^{-(t/\tau_{s,f})/\tau_{s,f}} \tag{2}
\]

\[
N_a = \min (N, F/5) \tag{3}
\]

\[
F_j = F - 5(j - 1), \quad j = 1, ..., N_a \tag{4}
\]

where \( p_m \) is the resulting time series of contraction force for a given muscle \( m \), \( n_s \) and \( n_f \) are the numbers of slow and fast fibers in a GMU, respectively, \( p_s \) and \( p_f \) are the slow and fast twitch responses (for the slow and fast fibers), respectively, and \( \tau_s \) and \( \tau_f \) are the respective time constants for the twitches.

Equation 1 defines \( p_m \) as a successive summation of twitches, considering the effect of each GMU. An infinite spike train is considered as an input, represented by the summation in \( k \). Each time a spike arrives, a twitch is triggered. This twitch is mathematically represented by equation 2.

Equation 3 defines \( N_a \) as the maximum number of active GMUs, considering a firing rate \( F \), and equation 4 defines the firing rate of recruited GMUs (rule of five).
For the twitch waveforms, the alpha function was used [8], which is described in equation 2. For the time constants, values of $\alpha_c = 35[ms]$ and $\gamma_f = 15[ms]$ were used for slow and fast twitches respectively [12] (see equation 2). At the same time, N=10 GMUs were considered for every simulation [12].

The simulation time was 5 seconds, discarding the first second, so the transient of the system was not considered in the statistical analysis. It is important to notice that each time a signal was generated, the mean value of $\alpha_m$ was different. To correctly estimate the mean value of $\alpha_m$, 40 simulations were considered for each frequency of fire.

![Figure 3: Example of muscle activation for the TA muscle, firing at 100[Hz]. Activation is around 0.5. For this case, 1[s] is simulated. It can be seen that in this case, it takes around 0.2[s] to get to stationary state.]

### 3. Results

The relationship between rate of fire and mean activation is first characterized. Although this was done for the TA muscle, the result is the same for LCA. The mean activation was computed by averaging all mean activations obtained at each fire rate, starting from 10[Hz], up to 300[Hz]. The result is illustrated in Figure 4. The range between 40[Hz] and 180[Hz] is known as dynamic range. In this range, the model has a linear behavior. Below this range, some GMUs start to become inactive, thus decreasing the effect on mean activation. On the other side, above 180[Hz] GMUs start to fire at their maximum capacity, reaching tetanization at around 200[Hz]. When tetanization is reached, activation saturates just below $\alpha_m = 1$.

The linear relationship provides a mapping between fire rate and mean activation, thus allowing an analysis of the effect of noise in the activation on models that consider this parameter as fixed. For the purpose of this study, the body-cover model (BCM) developed by Titze and Story [14] was considered. This low-dimensional model was chosen due to its simplicity, but also because it enables a natural connection between contraction of the cricothyroid and thyroarytenoid muscles and stiffness parameters. This dependence is ruled by a set of physiologically-based equations described also by Titze and Story [5].

#### 3.1. Effect of Neural Noise on the Body-Cover Model

The BCM model is controlled by a set of parameters that depend on the muscular activation of different muscles, like TA and LCA. For the following simulations, fire rates for TA and LCA are set at values in which mean activations are the same as fixed activations in the simulation without noise. This values are $a_{TA} = 0.25$ and LCA at $a_{LCA} = 0.5$. Also, fixed activation for the non-simulated muscle cricothyroid is set at $a_{CT} = 0.2$.

Figures 5a and 5b show that with a noisy simulation of TA activation, a variation in time can be incorporated in parameters of the body-cover model. Also, this are just two of the many parameters that are affected by this.

For a given combination of muscle activations, the BCM can enter on a self-oscillating state, or can start a damped oscillation. With this behavior in mind, a comparison between a constant and a noisy activation was made, so the effect of noise can be directly seen on the VF oscillation. Figure 6 illustrates this idea. Muscle activation for LCA is simulated, introducing noise in the model and showing how it reflects on a glottal area graph. As it can be seen, with a combination of activations of $a_{TA} = 0.25$, $a_{LCA} = 0.5$ and $a_{CT} = 0.2$, the model enters in a stable oscillation state. With noise however, this oscillating state becomes very irregular in its amplitude, but keeping its fundamental frequency relatively constant.

The muscle activation plot (MAP) that Titze and Story presented in their study [5] can be used to understand the results presented in Figure 6. Data points show region of self-sustained oscillation with the convergence rule, and for the given combination, oscillation is effectively happening. However, self-sustained oscillation state is limited by a narrow band around 0.5 for lateral cricoarytenoid activity. This implies that if noise is induced in LCA activation, then the BCM could continue to step in and out of the self-sustained oscillation state, which explains the irregular behavior in glottal area waveform.

Figure 7 shows a second comparison, in which noise in TA...
Figure 5: Effect of noise in TA muscle activation on vocal fold length (5a) and coupling constant (5b) of the BCM. Dashed line shows the value with no noise on the activation, considering the mean value of the noisy case.

Figure 6: Comparison between a simulation using the normal and noisy activation. Activation on cricothyroid (CT) and lateral cricoarytenoid (LCA) are set at 0.2 and 0.5, while TA had a mean value of 0.25.

The neural noise scheme for muscle activation presented in this study can be used to simulate a different set of muscles simultaneously. In future efforts, cricothyroid will be simulated, given its impact on fundamental frequency regulation [15]. A comprehensive sensitivity analysis will be performed to assess the effect of the neural noise in various configurations.

5. Acknowledgements

This work was funded by CONICYT grants FONDECYT 1151077, BASAL FB0008, and MEC 80150034, and the Ontario Ministry of Research and Innovation grant ER13-09-269.

6. References


