

VOICE SYNTHESIS USING THE GENERALIZED PRESSURE-CONTROLLED VALVE

Tamara Smyth and Alireza Fathi
School of Computing Science
Simon Fraser University

ABSTRACT

Vowel production in human speech depends both on the vocal tract shape, primarily establishing formant frequencies in the speech spectrum, and the vibration of vocal folds in the larynx, which function as a pressure-controlled valve regulating airflow into the vocal tract. This research explores the application of the generalized dynamic pressure-control valve model to the synthesis of voiced sounds in human speech. Where many vocal models employ a source filter paradigm, the dynamic model allows for feedback from the vocal tract to the glottal source, allowing for stronger influence of the vocal tract on the vibrations of the vocal folds.

1. INTRODUCTION

Many vocal systems, like reed-based musical instruments, use a pressure-controlled valve as the primary mechanical resonator for sound production. In both cases, an input pressure influences the valve's oscillation by creating a pressure difference across its surface. Any change in the up- and down-stream pressure of the valve will impact the behaviour of its oscillation by altering the overall driving force and causing the valve to open or close further.

Voiced sounds such as vowels in human speech production may be articulated using the velum, jaw, tongue and lips to shape the vocal tract, thus influencing formant frequencies in the waveform's spectrum. The speaker also controls the mechanical vibration of the vocal folds, a pressure-controlled valve which modulates airflow into the vocal tract and strongly influences the fundamental frequency, or pitch, of the sounding voice. In many cases, the source-filter model (Figure 1) is considered to be sufficiently accurate to simulate this interaction, as there is a weak coupling between the relatively massy vocal folds and the vocal tract. An improved synthesis is expected however, when taking the down-stream pressure of the valve (the pressure at the entrance to the vocal tract) into consideration when computing the force driving the valve oscillation and the flow through the valve channel.

More physically informed models are also available, many using digital waveguide synthesis and incorporating a series of two-port scattering junctions to simulate the varying cross-section of the vocal tract corresponding to the production of a particular vowel sound [1]. One- and two-mass spring models have been used for the vocal folds

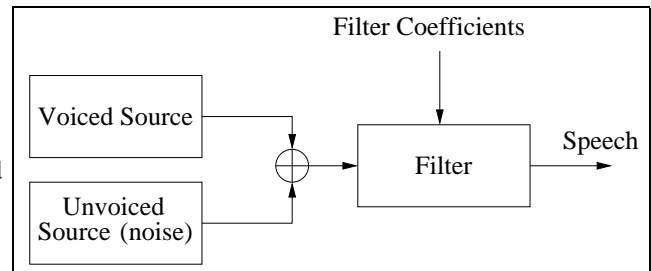


Figure 1. A source-filter (feed-forward) model for speech synthesis.

[2, 3, 4], with the two-mass model making it possible to capture a more idealized cycle of the vocal fold vibration, that is, with lower portion of the vocal folds leading the upper portion, creating a wave-like motion on the vocal fold surface [5].

In this work, we use the generalized pressure-control valve model to simulate the vocal tract, as it was shown to yield expected results for blown open, blown closed, and swinging door valve configurations [6]. The feathering incorporated in [7] is also expected to help capture an aspect of the wave-like motion of the vocal-fold displacement by smoothing the transition between an open and closed valve.

Though an accurate synthesis of the human voice requires a more complete implementation of the mouth and nose (see Section 3), we focus on the synthesis of vowel sounds by integrating the generalized valve, in a configuration and with parameters corresponding to the glottal source, within a waveguide model. In the final section, a software is presented, which provides a GUI for the control of the vocal tract model variables, as well as a graphical shape viewer, allowing the vocal tract shape to be viewed while listening to the corresponding vowel sounds.

2. MODELLING THE GLOTTAL SOURCE

2.1. Source-Filter Model

The vibration of the vocal folds produce the voiced sounds of speech, and in particular, that of vowel sounds. Many techniques for modeling vowels use a source-filter model, which incorporates a feed forward filter (i.e. one that does not feed back to the source), creating a situation where the resonance of the vocal tract has no impact on the os-

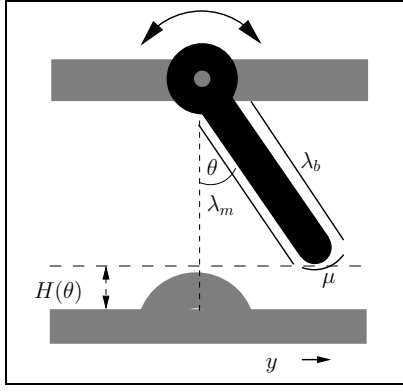


Figure 2. The blown open configuration of the generalized valve model, showing geometric parameters λ_{sg} , the length of the valve that sees the sub-glottal pressure, λ_d , the length of the valve that sees the valve's downstream pressure, and μ , the length of the valve that sees the flow. Changing these parameters will change the corresponding component forces of the overall driving force in (2).

cillation of the source. That is, the source, or a glottal excitation, is typically simulated using a filtered periodic impulse train, creating a *voiced source* with a fundamental frequency, plus any additional unvoiced sources such as noise. The source is then passed through a filter with coefficients set according to the vowel being produced and/or the corresponding vocal tract shape (see Figure 1). For many cases, this model is considered to be sufficiently accurate as there is a relatively weak coupling between the massy vocal folds and the vocal tract.

2.2. The generalized pressure-controlled valve

Using a dynamic model, like the generalized pressure control valve, allows for a stronger influence of the filter on the source. Unlike the traditional source-filter, where signal flow is strictly from left to right (see Figure 1), the dynamic valve model has feedback resulting from the use of the down-stream pressure p_d when computing the valve's time-evolving variables, and in particular, the valve displacement x and the volume flow U through the valve channel.

The generalized valve model was first introduced in [6], providing a configurable model of a pressure controlled valve, allowing the user to design their own virtual reed type, simply by setting the model's parameters. Here, we place the pressure-controlled valve model in the context of the human vocal tract for an improved synthesis of voiced sounds such as vowels.

The model's parameters are continuously variable, and may be configured to produce blown closed, blown open, and the symmetric "swinging door" models, as well as to set the valve geometry. Though the fleshy nature of the human vocal folds prevent it from vibrating with as predictable behaviour as the more rigid woodwind reed, the dominant motion of the valve is considered to be blown open, that is, a sub-glottal pressure increase will cause the

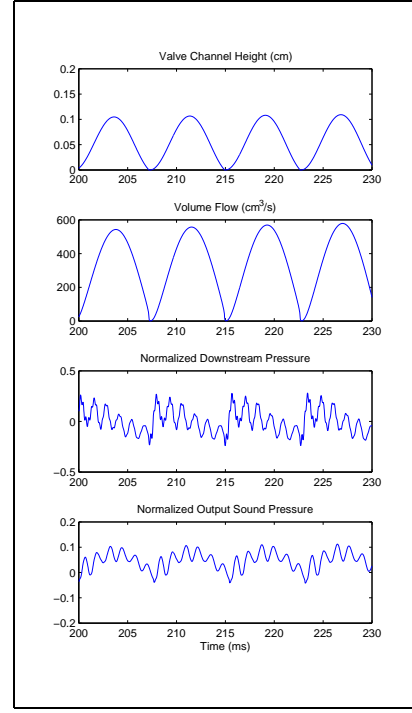


Figure 3. Simulation results for the vowel 'ah', showing the time evolution of the valve channel height, the volume flow through the channel, and the normalized pressure at the entrance to the vocal tract as well as at the mouth.

valve to open further, while an increase in down-stream pressure will force the valve to close further (see Figure 2). It may be possible to create more complex oscillatory behaviour by having two valves with different configurations placed in cascade, but this is left for future work.

It is by considering both the force F driving the reed, and the air flowing through the reed U , that the dependence of valve's oscillation on the down-stream pressure is apparent. The displacement of the valve may be approximated by the second-order differential equation

$$m \frac{d^2 x}{dt^2} + m 2\gamma \frac{dx}{dt} + k(x - x_0) = F, \quad (1)$$

where m is the effective mass of the reed, γ is the damping coefficient, k is the stiffness of the reed. The overall driving force is equal to the sum of all forces acting on the reed,

$$F = F_m + F_b + F_U, \quad (2)$$

where the force F_m acts on the valve surface area seen by the sub-glottal pressure p_{sg} , the force F_b acts on the valve surface area seen by the down-stream pressure p_d , and the force F_U is applied by the flow U , forcing the valve open. Example parameter values are seen in Table 1.

The differential equation governing air flow through the valve, fully derived in [7], is also dependent on down-stream pressure (that is, the pressure difference across the valve) and is given by

$$\frac{dU}{dt} = (p_{sg} - p_d) \frac{A(t_0)}{\mu\rho} - \frac{U(t_0)^2}{2\mu A(t_0) + U(t_0)T}, \quad (3)$$

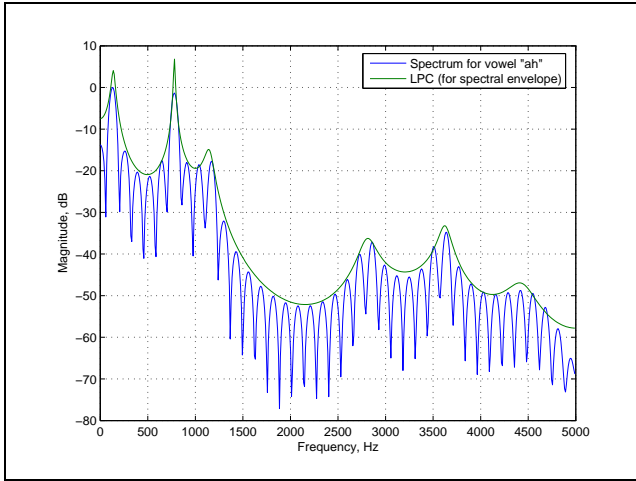


Figure 4. Spectrum and spectral envelope (extracted using linear predictive coding (LPC) of order 16) for the vowel ‘ah’.

Quantity	Numeric value
Valve type (oscillation constraint)	blown open
Radius of exhaust pipe a	0.014 m
Quality factor (Q) (valve damping)	9.0
Mass (m)	0.00028 kg
Width of the valve (w)	0.0049 m
Length of the valve (l)	0.014 m
Thickness of the valve (d)	0.003 m
Equilibrium position (x_0)	0.0002 m

Table 1. Synthesis control parameters and example values (as suggested by [2]).

where $A(t)$ is the cross sectional area of the valve channel, and μ is the length of valve that sees the flow (see Figure 2).

The oscillating vocal folds may therefore be modeled digitally by obtaining a value every sample period for the displacement of the valve x , the flow U , and the pressure at the entrance to the vocal tract p_d , in response to an applied sub-glottal pressure p_{sg} . Figure 4 shows the time-evolution of these parameters in response to an applied sub-glottal pressure $p_{sg} = 800$ Pa, and using a vocal tract with the shape corresponding to the vowel sound ‘ah’ (top of Figure 6).

3. THE VOCAL TRACT SIMULATION

The pressure at the entrance to the vocal tract is obtained from the waveguide model which simulates the delay in the vocal tract, as well as the scattering which occurs along its length due to the varying cross section. The characteristic impedance at each of the ports in the sequence of two-port scattering junctions (see Figure 5), is derived from a vocal tract area function corresponding to the vowel being synthesized (see Figure 6). With known impedances on either side of the junction, the junction pressure is deter-

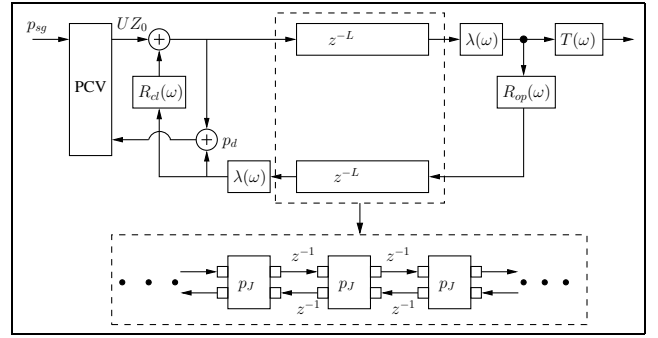


Figure 5. The model showing feedback between the waveguide delay element and the pressure-controlled valve model. The delay element may be implemented as a one-dimensional waveguide for a simple cylinder, or as a series of two-port junctions for a tube—or vocal tract—with varying cross section.

mined by

$$p_j = \frac{2 \sum_{i=1}^N \Gamma_i p_i^+}{\sum_{i=1}^N \Gamma_i} \quad (4)$$

where $\Gamma_i = 1/Z_i$ and $N = 2$ for a two-port junction, and the output pressure at any port is simply the difference in the junction pressure and the incoming pressure on that port, $p_i^- = p_j - p_i^+$.

In addition to the vocal tract shape simulation as a piecewise connection of several cylindrical tubes connected by two-port scattering junctions (see Figure 5) [8], there is a three-port scattering junction connecting the glottal tube to oral and nasal tubes (where the junction pressure is obtained using (4) but with $N = 3$) [1]. The effects of this addition to the overall vocal tract shape may be controlled using the GUI seen in Figure 7. The slider labeled *Mouth Reflection* allows for control of the mouth opening, while the slider labeled *Nose Reflection* allows for control of the nasal quality of the produced sounds (these corresponds to the controls in the SPASM software [9]).

The graphical user interface for the vocal tract simulation is implemented in OpenGL, using a graphical visualizer that allows for viewing of the vocal tract shape while listening to the corresponding vowel sounds. The shapes were obtained from medical resonance imaging (MRI) [10, 11] data, from which vocal tract area functions were extracted and used to generate the images seen in Figure 6.

4. CONCLUSION

In this work we incorporate the generalized pressure control valve, previously seen in the context of wind instrument, to simulated voiced sounds of human speech. The valve configuration and anatomical parameters for a the vocal folds are given, with results showing a successful simulation of vowel sounds. A graphical user interface is provided, allowing for control of the vocal tract, as well as providing a visualization of the vocal track shape corresponding to the sound being produced.

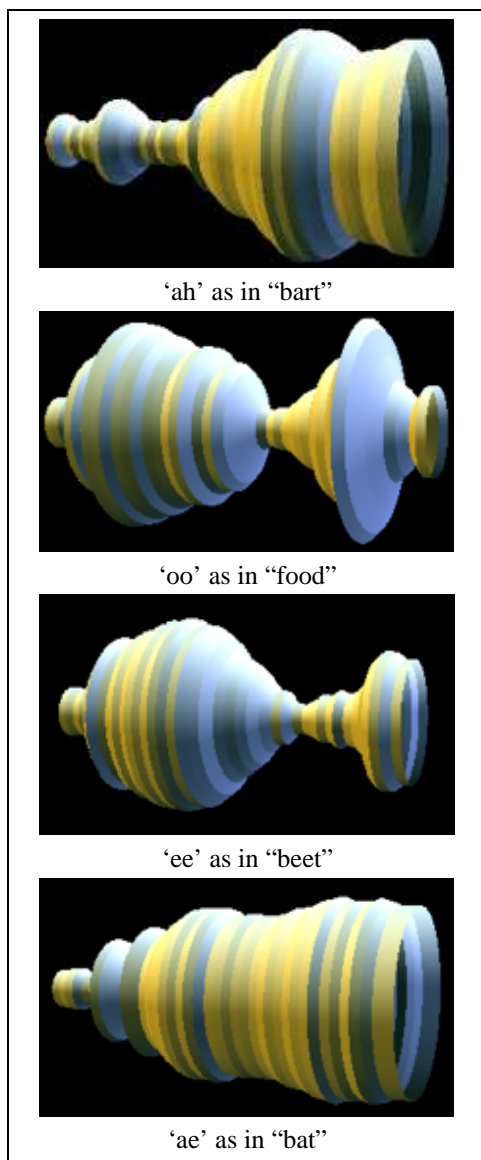


Figure 6. Vocal tract shapes corresponding to vowel sounds, ‘ah’, ‘oo’, ‘ee’, and ‘ae’, respectively, generated by the GUI’s graphical visualizer.

5. REFERENCES

- [1] Perry R. Cook, *Identification of Control Parameters in an Articulatory Vocal Tract Model, with Applications to the Synthesis of Singing*, Ph.D. thesis, Stanford University, Stanford, California, December 1990.
- [2] Seiji Adachi and Jason Yu, “Two-dimensional model of vocal fold vibration for sound synthesis of voice and soprano singing,” *Journal of the Acoustical Society of America*, vol. 117, no. 5, pp. 3213–3224, May 2005.
- [3] Xavier Rodet, “One and two mass model oscillations for voice and instruments,” in *Proceedings of ICMC 1995*, Banff, Canada, 1995, International Computer Music Conference.

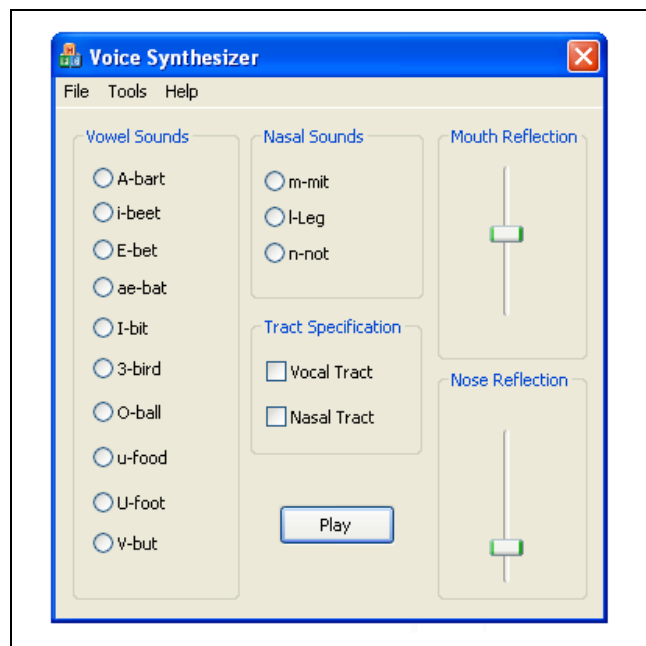


Figure 7. A graphical user interface for controlling the valve shapes.

- [4] Carlo Drioli, *Voice coding by means of physically-based models*, Ph.D. thesis, Dept. of Information Engineering, University of Padova, 2002.
- [5] Brad Hudson Story, “An overview of the physiology, physics and modeling of the sound source for vowels,” *Acoustical Science and Technology*, vol. 23, no. 4, pp. 195–206, 2002.
- [6] Tamara Smyth, Jonathan Abel, and Julius O. Smith, “A generalized parametric reed model for virtual musical instruments,” in *Proceedings of ICMC 2005*, Barcelona, Spain, September 2005, International Computer Music Conference, pp. 347–350.
- [7] Tamara Smyth, Jonathan Abel, and Julius O. Smith, “The feathered clarinet reed,” in *Proceedings of the International Conference on Digital Audio Effects (DAFx’04)*, Naples, Italy, October 2004, pp. 95–100.
- [8] Julius O. Smith, *Digital Waveguide Modeling of Musical Instruments*, ccrma.stanford.edu/~jos/waveguide/, 2003.
- [9] Perry R. Cook, “Spasm, a real-time vocal tract physical model controller; and singer, the companion software synthesis system,” *Computer Music Journal*, pp. 30–44, 1993.
- [10] Gunnar Fant, *Acoustic Theory of Speech Production*, Mouton, The Hague, 1960.
- [11] Brad Hudson Story, *Physiologically-based Speech Simulation using an Enhanced Wave-Reflection Model of the Vocal Tract*, Ph.D. thesis, University of Iowa, May 1995.