The mathematics of singing:

An overview of mathematical models of the singing voice

Elizabeth Reid
Department of Computer Science
University of British Columbia
reid42@cs.ubc.ca

Jennifer Lang
Department of Music
University of Saskatchewan
jennifer.lang@usask.ca

Steven Rayan
quanTA/Department of Mathematics & Statistics
University of Saskatchewan
rayan@math.usask.ca

University of Saskatchewan

Motivation

Singing is a versatile, powerful means of artistic expression, yet physically singing is quite a complex process. Learning about some of the mathematical principles that explain why the human voice sounds the way it does can help us understand our own voices better, add effects to sound recordings, or even synthesize entirely new voices.

For example, voice analysis allows us to study the intricacies of sound production, create visualization tools, and has implications for vocal pedagogy and the study of voice pathologies. On the other hand, voice synthesis facilitates the creation of entirely original music, and may be particularly useful for composers. Analysis and synthesis can even be combined to modify existing voice parts or create special effects.

Model Types

There are two main types of voice models: ***physical*** and ***spectral***.

Physical Models represent the anatomical interactions that occur during sound production. These models require a deep understanding of vocal anatomy and physics to develop and use effectively.

Spectral Models represent the properties of sound waves after they have left the mouth. These models do not rely on sophisticated knowledge of vocal anatomy, but instead rely on the properties of sound waves and the principles of sound wave interaction.

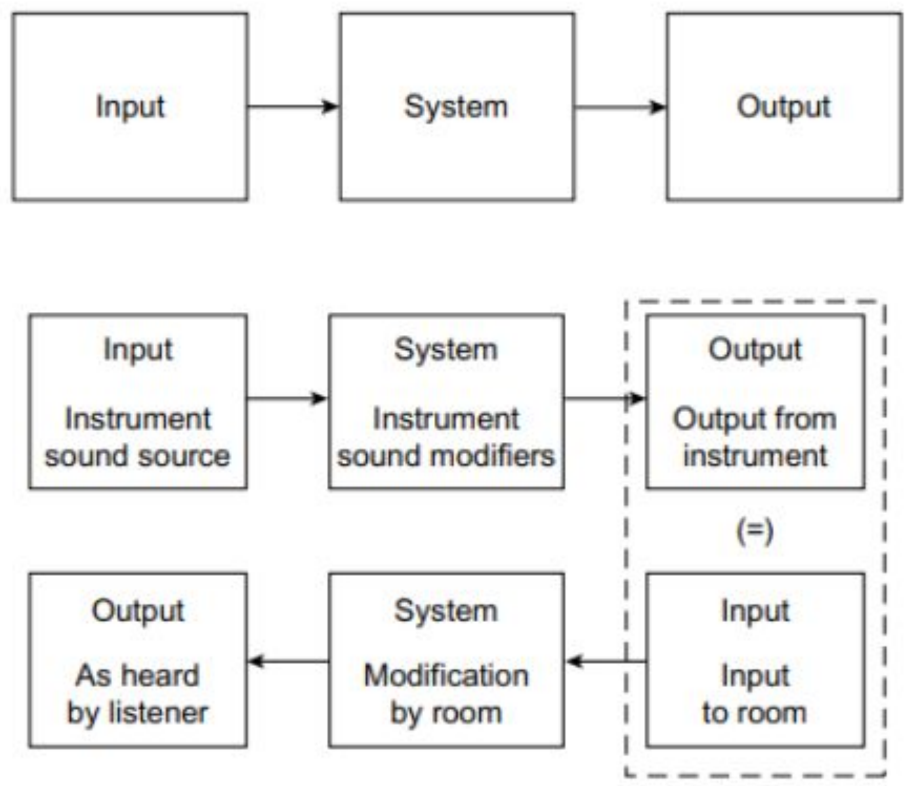


Fig. 1: A diagram of the musical sound production system, reproduced from Howard and Angus [3].

Physical Models

Singing requires the interaction of breath, oscillators, resonators, and articulators. Physical models typically focus on two parts of sound production: the sound source, and the resonating space, which are modelled independently and later combined together.

The sound source: the interaction of air moving through the vocal folds.

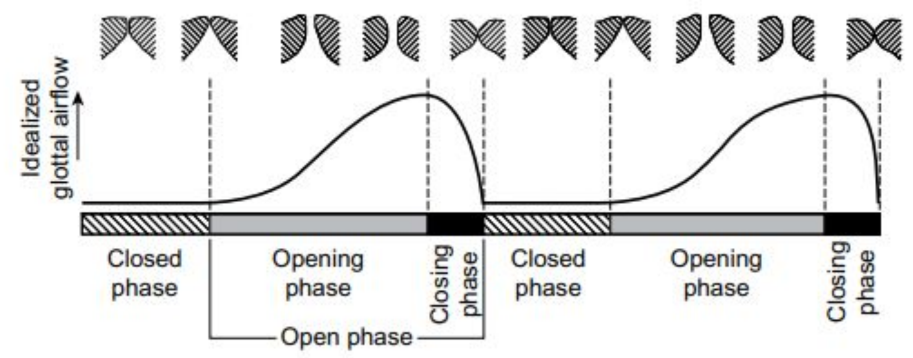


Fig. 2: A simplified visualization of vocal fold movement, reproduced from Howard and Angus [3].

The vocal folds are represented as a number of individual masses: forces act on the masses, which determine the mass positions over time.

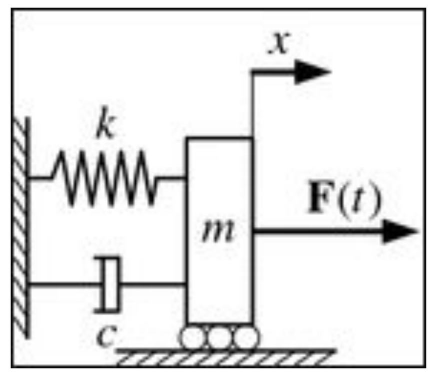


Fig. 3: A diagram of a one mass vocal source model, reproduced from Simon et al [5].

An example of a simple vocal fold model proposed by Flanagan [2] is given as:

$$m\ddot{x} + c\dot{x} + kx = \frac{1}{2} l d \left(P_s - 0.935\frac{\rho v^2}{L^2}\right)$$

where v is the velocity of air flow, l corresponds to the length of the vocal fold masses, d corresponds to the vocal cord thickness, P_s is the subglottal pressure, ρ is the air density, and L is the glottal area.

The resonating space: the vocal tract resonators enhance some overtones and dampen others, giving the human voice its characteristic sound. Vocal tract models aim to simply the wave equation:

$$\nabla^2\psi = \frac{1}{c^2}\frac{\partial^2\psi}{\partial t^2}$$

by representing the vocal tract as a series of cylinders or cones. Forward and backward travelling waves are then calculated for each cylinder by the equation [6]:

$$\psi_k(x,t) = \psi_k^+(x-ct) + \psi_k^-(x+ct)$$

where ψ_k^+ is the forward component of sound pressure in section k, ψ_k^- is the backward component, c is the speed of sound, and t is time.

Spectral Models

There are many important properties of sound waves, including intensity, directivity, and sound scattering, but for spectral models, one of the most important concepts arises from Fourier analysis.

In particular, any waveform can be built up from an appropriate set of sine waves, and any waveform can be deconstructed into a set of sine waves. These individual sine waves are the partials or harmonics of the waveform, and we can break up a waveform into harmonic parts, deal with each separately in a model, and recombine the results.

The discrete Fourier transform and inverse discrete Fourier transform are given as follows [4]:

$$F(k) = \frac{1}{N}\sum_{n=0}^{N-1} f(n)e^{\frac{-2\pi ikn}{N}}$$

where f(n) is the nth sample of the wave f (i.e., f(n) is the magnitude of f at the time n / f_s, where f_s is the sampling frequency of f), N is the total number of samples, and 0 ≤ k < N. The inverse transform is given by:

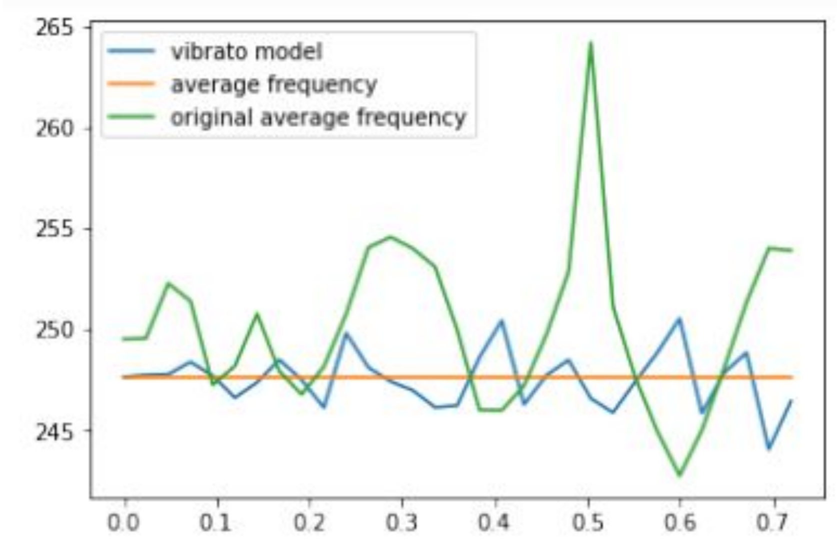
$$f(n) = \sum_{k=0}^{N-1} F(k)e^{\frac{2\pi ikn}{N}}$$

where 0 ≤ n < N.

Some types of spectral models rely on a similar method, called additive synthesis, which requires the frequency, amplitude, and phase of each sinusoidal component to be known. If that is the case, then the original wave f can be re-synthesized using the following equation [4]:

$$f(n) = \sum_{i=1}^{P} a_i(t) sin(nf_i(t) + \varphi_i(t))$$

where N is the length of the wave, 0 ≤ n < N, P is the number of component sinusoids, a_i, f_i, and φ_i are the amplitude, frequency, and phase of the ith sinusoid, t = n / f_s, and f_s is the sampling frequency of f. The example below shows how an additive synthesis model, first derived by Diaz [1], can be used to resynthesize vibrato. Note that the reconstruction is imperfect, likely due to the variation of vibrato extent and rate that occurs over the sample.



Discussion

Physical models:
- Require detailed knowledge of vocal anatomy and physics
- Involve solving (sometimes complex) differential equations
- Are very good at modelling specific vocal fold properties
- Don't easily generalize to different singers if specific anatomical measurements are not known

Spectral models:
- Require an understanding of basic sound properties, and are generally simpler
- Much more generalizable across different singers
- Especially good for analysis and re-synthesis
- Aren't well suited for certain representational tasks (e.g., complex voice pathologies)

Both modelling types can be more or less complex, and each is ideal for different types of modelling tasks.

Limitations and Future Work

- Machine learning approaches might also be useful to consider and compare
- Worth exploring how these models handle professional vs. amateur singers, as well as different styles
- Can these models be adapted to handle multiple singers?
- Can understanding models like these improve a singer's understanding of their own voice?

Acknowledgments

I would like to thank Dr. Jennifer Lang and Dr. Steven Rayan for their insightful supervision of this project.

References

[1] J. A. Diaz, "A mathematical model of singers' vibrato based on waveform analysis," Doctoral Dissertation, Acoustical Society of America, 1998. [Online]. Available: https://asa.scitation.org/doi/abs/10.1121/1.427809

[2] J. L. Flanagan, "Source-System Interaction in the Vocal Tract," *Annals of the New York Academy of Sciences*, vol. 155, no. 1, pp. 9–17, 1968, doi: https://doi.org/10.1111/j.1749-6632.1968.tb56744.x.

[3] D. M. Howard and J. A. S. Angus, "Acoustic Model for Musical Instruments," in *Acoustics and Psychoacoustics*, 4th Edition., New York: Routledge, 2009, pp. 167–229.

[4] K. Lomax, "The Analysis and Synthesis of the Singing Voice," Doctoral Dissertation, University of Oxford, 1997.

[5] A. Simón Mata *et al.*, "Vibrations in Systems with One Degree of Freedom," in *Fundamentals of Machine Theory and Mechanisms*, A. Simón Mata, A. Bataller Torras, J. A. Cabrera Carrillo, F. Ezquerro Juanco, A. J. Guerra Fernández, F. Nadal Martínez, and A. Ortiz Fernández, Eds. Cham: Springer International Publishing, 2016, pp. 249–277. Doi: 10.1007/978-3-319-31970-4_7.

[6] B. Story, "Physiologically-Based Speech Simulation Using an Enhanced Wave-Reflection Model of the Vocal Tract," Doctoral Dissertation, University of Iowa, 1995.