UNIVERSIDAD
NACIONAL
DE COLOMBIA

Design, development, and evaluation of a computational physical model to generate synthetic birdsongs from recorded samples

Bachelor Thesis

Sebastian Aguilera
Novoa

birdsongs

Bachelor dissertation, National University of Colombia

github.com/saguileran/birdsongs

This thesis was done under the supervision of professor Francisco Gómez Jaramillo, Datalab group leader. It is a requirement to obtain the degree in physics.

First release, December 2022

Diseño, desarrollo y evaluación de un modelo físico-computacional generador de cantos de aves sintéticos a partir de cantos grabados.

Sebastian Aguilera Novoa

Universidad Nacional de Colombia
Facultad de Física, Departamento de Ciencias Naturales.
Bogotá, Colombia
2022

Design, development, and evaluation of a computational-physical model to generate synthetic birdsongs from recorded samples

Sebastian Aguilera Novoa

Degree work presented as a partial requirement for the PHYSICS degree:
Bachelors in Physics

Director:
Francisco Gómez Jaramillo Ph.D.

National University of Colombia
Physics Faculty, Exact Science Department.
Bogotá, Colombia
2022

In honor of my mother, because she was the giant shoulders that allowed me to be a scientist.

"Education is the most powerful weapon you can use to change the world."

Nelson Mandela

Acknowledgment

I want to thanks all the teaching staff of the Universidad Nacional de Colombia, who guided me to achieve my dreams. Professor Francisco Gomez for being a great educator and teach me the tools and knowledge to carry out this work, Professor Jorge Ruiz for teaching me all the ideas of numerical optimization; to Juan Ulloa, for our discussion and his advices that guide me to formulate and define the problem presented; and Professor Gabo Mindlin for always answering my questions about the physics of the model and give me great advices.

I also want to thank all the my classmates and friends to support me in this adventure: to Jonnatan, for his life teachings and for showing me the meaning and importance of friendship; to Sami, for sharing and brightening up several years of my life with their company letting me know what it is to be loved; to all the Miguels with whom I shared, discussed and they teach me many things; to Raul for being the best high school professor and for always giving me encouragement to study; and of course, to all the Sebastians I have met along this adventure and who make me love my name. I would like to thank all those who get involved in this adventure and share their time with me, I will keep you in my mind and my heart.

Finally, I would also like to thank my father, for letting me be, and my brother, because his love for music went beyond the walls and splashed me, to both of them thanks for always being by my side and supporting my ideas and projects. And a special thank to my mother, for giving me life and the curiosity to live it, for supporting my education and showing me the beauty of life, because without her none of these will be possible.

Mother, thank you very much for making my dreams possible.

Abstract

Any Colombian has probably heard a birdsongs at one time or another. In fact, Colombia is the country with the largest bird population in the world playing an important role in the conservation of bird, since the richness of bird in an ecosystem will give information about the loss biodiversity and the sustainable of the ecosystem. As a physicist, I have been interested in how to use physical models to generate realistic data, especially in acoustic physics. I have studied how to simulate the production of sound by means of the means of the acoustics of musical instruments and the sound production of birds. However, musical acoustics are widely researched, as well the classification and identification of birds using their birdsongs, but the production of birdsongs is another matter. The best and most complete physical model that explains the production of sound in birds is the **motor gestures**[**?**] model, developed two decades ago i the Dynamical Systems Lab by professor G. Mindlin, which uses nonlinear ordinary differential equations to model the organs involved in the production of sound in birds: **syrinx**[1], trachea, glottis, oro-oesophageal cavity (OEC), and beak. For his purpose, a control problem of two parameters, air sac pressure from the bird bronchi and labial tension, is formulated such that it generates a synthetic sample from simple paths of the parameters space and recorded birdsongs.

The present work designs, develops and evaluates a python packing for the motor gestures model. Using current tools of signal processing, numerical optimization, and object programming oriented, the model is successfully implemented making it fast of reproduce and easier develop. With the presented model implementation, the sound production of birds can be broadly studied: the parameters space and their impact on the synthetic birdsong. This implementation makes it possible to create several samples of synthetic birdsongs, with a few python command lines, and compare them through their spectral characteristics. Teherefore, the model allows us to create as many not complex birdsongs, with a fundamental frequency well defined, as we want and by modifying the path in the space parameters (motor gesture). In addition, future applications will allow us to identify and classify the bird by its motor gesture.

The performance of the model is evaluated mainly with the birdsong syllables of the colored sparrow (zonotrichia capensis or copeton), since this is a familiar bird in Colombia, and its generalization is tested with other oscine birds[2]bird in the Passeri suborder are called oscines or birdsongs: Mimus gilvus, Euphonia laniirostris, and Rhinocryptidae. The results show that the model generates comparable birdsong syllables when the input audio is of high sound quality, low nose level, and the pitch is well computed. in other cases the model does not generate a comparable birdsong and may give a diverse birdsong.

[1] Main character for the sound production. A tissue that vibrates causing the air oscillates and modulate a sound pressure input to the trachea

Key Words: birdsong, numerical-optimization, signal-processing.

Contents

Acknowledgment iv

1 Introduction 2

2 Literature Review 6
- 2.1 Waves . . . 6
  - 2.1.1 Sound Waves . . . 7
- 2.2 Signal Processing . . . 7
  - 2.2.1 Time Domain . . . 8
  - 2.2.2 Spectral Analysis . . . 9
  - 2.2.3 Acoustic Indexes . . . 13
- 2.3 Dynamical Systems . . . 14
- 2.4 Bifurcation Theory . . . 15
  - 2.4.1 Types of Equilibria . . . 16
  - 2.4.2 Andronov-Hopf . . . 20
  - 2.4.3 Saddle-Node . . . 21
  - 2.4.4 Limit Cycle . . . 22
  - 2.4.5 Bogdanov–Takens (BT) Bifurcation . . . 22
- 2.5 Numerical Optimization . . . 23
  - 2.5.1 Introduction . . . 24
  - 2.5.2 Inverse Problem . . . 27
  - 2.5.3 Non-Linear Optimization . . . 29

3 The Problem and its Background 30
- 3.1 Birdsong . . . 30
  - 3.1.1 Bird Calls and Syllables . . . 30
- 3.2 Sound Production of a Birdsong . . . 32
  - 3.2.1 Syrinx . . . 33
  - 3.2.2 Physics of a Birdsong . . . 35
- 3.3 Motor Gestures . . . 37
  - 3.3.1 Timeline . . . 38
  - 3.3.2 Current Model . . . 42
  - 3.3.3 State of Art . . . 46

3.4 Automatizing . . . 46
3.4.1 Optimization Problem . . . 47

4 Methods and Methodology **48**
4.1 Programming Object Oriented (POO) . . . 48
4.1.1 Python Objects . . . 49
4.2 Model Implementation . . . 49
4.2.1 Audio Processing . . . 49
4.2.2 Methodology . . . 49
4.2.3 Numerical Optimization . . . 50

5 Results and Discussion **53**
5.1 Model Implementation . . . 53
5.1.1 Motor Gestures . . . 53
5.1.2 Syllable . . . 57
5.1.3 Chunck . . . 59
5.1.4 Birdsong . . . 60
5.2 Evaluation . . . 63
5.2.1 Consistency . . . 63
5.2.2 Uniqueness . . . 64
5.2.3 Generalization . . . 64

6 Conclusions **66**
6.1 Conclusions . . . 66
6.2 Boundaries . . . 66
6.3 Future Works . . . 66

1 Introduction

At present, the world population has extremely increased to 8 billions of persons generating more natural resource consumption and placing extraordinary demands on agriculture and ecosystem services. As specie, we have developed amazing technology: telescopes (Webb Space Telescope) able to watch the universe through great solution, powerful and immense particle accelerator (LHC) where many particles have been discovered and studied, or even extraordinarily computers that make possible to simulate an entire gene of DNA using a billion-atom bio-molecules int the Los Alamos National Laboratory. Nevertheless, our natural resources demanding has reached a unsustainable point where we consume much more than we produce leading the malnourished of an important part to the whole population, more than a billion of persons. Hence, as a scientist we have to keep in mind solutions for a conserve the planet and ecosystems [**?**].

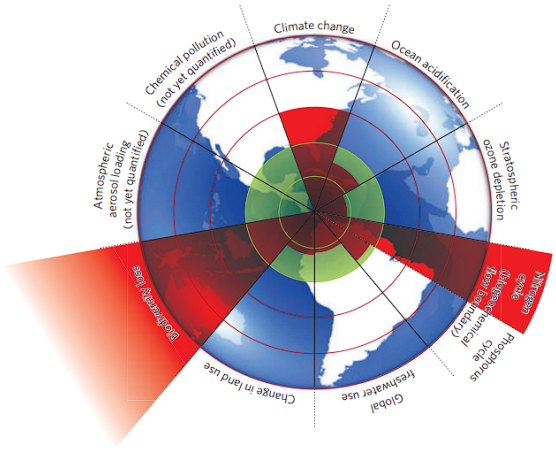


Figure 1-1: Nine planetary systems are studied in order to make a safe operating space for humanity. The green shaded region represents the safe operating system space while the red represents an estimate of the current position for each variable showing the safe values have already been exceeded.[**?**]

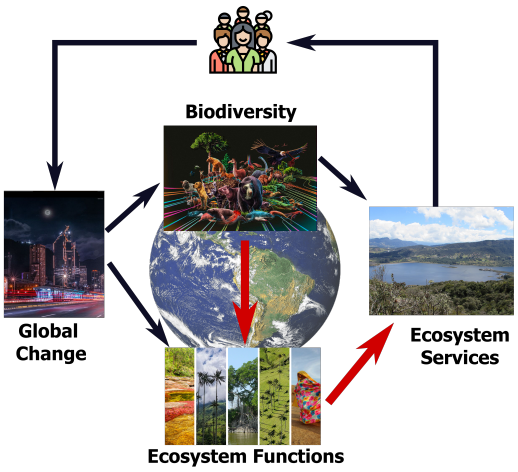


Figure 1-2: Scheme of world interactions with humans and world, ecosystem functions and services for living beings. [**?**]

But, how can we preserve the ecosystems? what are the boundaries that human beings must keep to avoid unacceptable environmental changes? These questions and many other have being

answered with the goal to identify and quantify what are the principal systems and events that affect the most the world environment. A naive idea may be that the most important factor of environment loss is the climate change but research have shown that even if it does impact the ecosystem, it is not the main cause of the loss. Good indexes of environmental change describing may include the biodiversity loss and human interference with the nitrogen cycle **1-1**, furthermore they may include some food security and environmental goals to have a sustainable process.

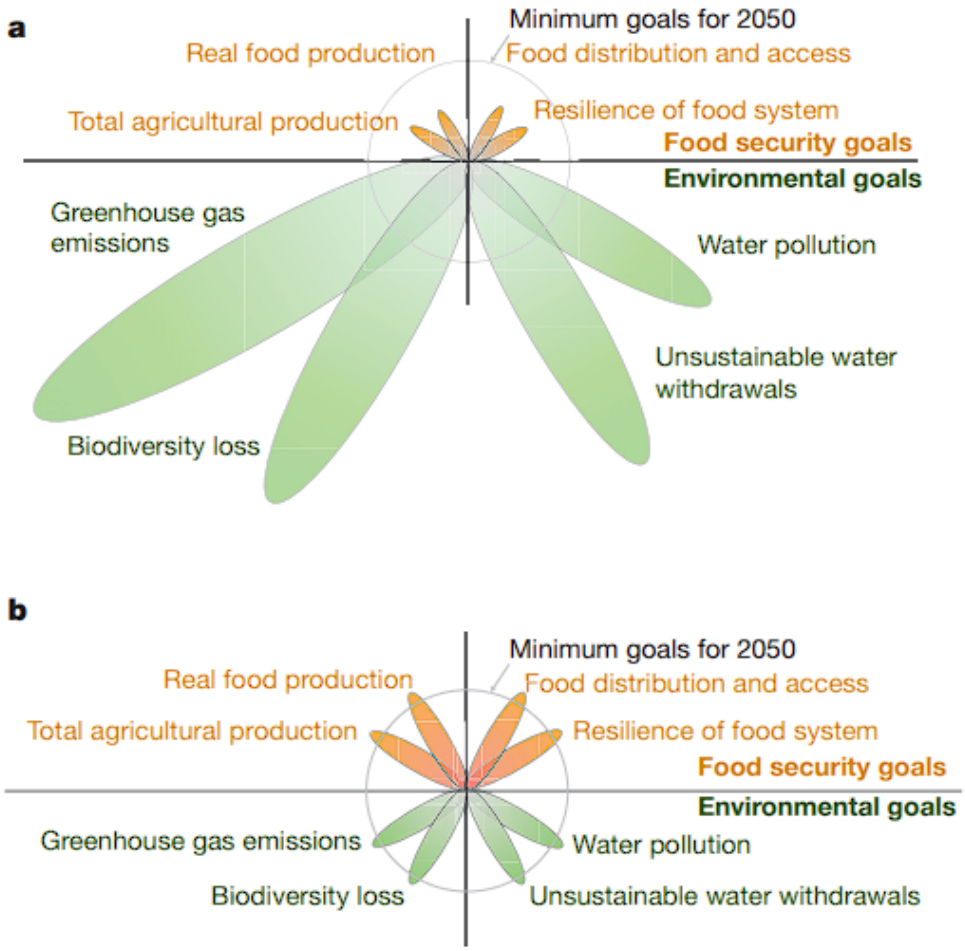


Figure 1-3: Meeting goals for food security and environmental sustainability by 2050. . Qualitatively illustration of a subset of the goals agriculture must meet in the coming decades. At the top, we outline four key food security goals: increasing total agricultural production, increasing the supply of food (recognizing that agricultural yields are not always equivalent to food), improving the distribution of and access to food, and increasing the resilience of the whole food system. At the bottom, we illustrate four key environmental goals agriculture must also meet: reducing greenhouse gas emissions from agriculture and land use, reducing biodiversity loss, phasing out unsustainable water withdrawals, and curtailing air and water pollution from agriculture. [**?**]

Hence, the ecosystems preservation should include the birds population conservation, a good ecosystem richness index, since birds are anywhere and have an important rule for the ecosystems. Birds offers many services to society: pollination, spread seeds, decompostion, control pests, and even their pop is a fertilizer, so birds transform entire landscapes making them richer in biodiversity and if their missing will causes big impacts to the ecosystem process, as may be the case of Colombia that is ranked as the country with the largest variety of birds in the world. But this it not all, birds have also been studied to understand the human brain and learning process, and even human sound production. Birds have many similarities with humans: they learn to sing from a tutor [**?**] (as humans learns to speak by hearing and repeating other humans vocalizations) and then process and reinforce this information by signing while they are sleeping[1] [**?**, **?**] showing that even when humans learning is more complicated birds can be used to explore and test human neurology; other articles explore the analogy between birdsong syllables and music theory reveling music-like dynamic structure in songbird rhythms [**?**]; among others.

In the process of studying similarities between birds an humans, computer scientist have used the current audio signal processing and machine learning ideas to contribute to the birds conservation problem. Those research have studied how recognize, identify, and classify birds by their birdsong or photography. Awesome research and tools have been created making possible to identity birds with just an smartphone, as does the Merlin Sound ID app develop by the Cornell Lab of Ornithology. However, in spite of the huge quantity of birdsongs recorded just a few of them are individual birdsongs and have great sound quality, with low noise level and high syllables spectral resolution, which leads to a scarcity of individual birdsongs audio records. As physicist, I have been always interested in how sound is produced and how physical models and computer tools can be used to generate realistic data. In this way, I started to study the physics of sound production and how is it simulate, mostly in music instruments, with the propose to generate synthetic sounds. Since music have been widely investigated, there are many simulations available on internet that simulate different instruments, in fact, you can find a huge quantity of sound simulations and even vocal fold simulation, human sound production.

This work present an study and packing of the motor gestures physical model, a python package called birdsongs.

[1] their brain sends electrical signals to the muscles, without sound generation, making a singing learning reinforcement by repeating the neurons activities while they are sleeping

The necessary concepts and equations are describe in the literature review chapter 2. Next, the problem of birdsongs production and its background is explain in the chapter 3 where at the end an automatization is proposed and discussed using optimization theory and algorithms. The programming model implementation and automatizing, solution of a minimization problem, is described and explain in the chapter 4 followed by the results and discussed, chapter 5. Finally, some conclusion are present where the package bounds and future works are specified.

Figure 1-4: birdsongs: a Python package to analyze, visualize and generate synthetic birdsongs using the *motor gestures model*.

2 Literature Review

This chapter contains concepts and equations necessary for the work presented. The chapter begins by defining a sound wave in terms of physics and then in terms of signal processing, some spectral features are defined in both theory and algorithm. This is followed by an introduction to dynamical systems and bifurcation theory describing the type of bifurcations. The final section of this chapter is concern with defining nonlinear optimization and describing the necessary concepts.

2.1 Waves

In fluid dynamics a wave is defined as the propagation of a perturbation of one or more physical variables, usually periodic and some times having a well defined frequency (periodic motions), which is created by the vibration of an object, called source of vibration. When the wave is in motion is called a *traveling wave*, while a *stationary wave* is a wave such that it remains in a constant position. Waves can overlap, they can combine to make constructive or destructive interference.

When the wave moves in a medium it is a *mechanical wave*, while when it does not require a medium to travel through it is an *electromagnetic wave*, propagating through vacuum (like light). Another classification has to do with the direction of the motion and vibration. A *longitudinal wave* has the same direction of vibration as the direction traveling, while a *transverse wave* moves perpendicular to the direction of oscillation.

Using differential calculus the most general equation that a wave, regardless of the type or nature of the wave, satisfies is a linear second-order partial differential equation which, in the most general case, can be a forcing[1] equation

$$\frac{1}{v^2}\frac{\partial^2 x(t,\vec{r})}{\partial^2 t^2} - \nabla^2 x(t,\vec{r}) = f(t,\vec{r}) \tag{2-1}$$

where $v = ||\vec{v}||$ is the magnitude of the sound speed, $x(t,\vec{r})$ the amplitude of the wave, and $\vec{r}$ the vector position. A particular solution is any *traveling wave*, a space and time depending function satisfying $x(x,t) = x(\vec{r} \pm \vec{c}t)$, or even a linear combination of them, since the wave equation

[1] the right-hand side represents a non-homogeneity function that acts as source of oscillations ($f(t,\vec{r})$), it is often called the forcing or driving function

is a linear hence satisfies the superposition principle. [**?**]. The backward (or leftward) wave is $x(\vec{r}+\vec{ct})$, while $x(\vec{r}-\vec{ct})$ represents a forward (or rightward) wave.
The energy transport by the wave is related to their frequency and amplitude.

2.1.1 Sound Waves

An interesting case is the **sound waves**. When the disturbing variable that propagates is the pressure that generates movement of the particle and the variation of the local pressure, it is due to particle-particle interactions, being a mechanical wave requires a medium to travel. The sound travels longitudinally in the medium, air or water, by displacing the air particles from its equilibrium position, they exerts a push or pull on their neighborhoods causing them to be displaced from their equilibrium position, Figure 3-9.

Hence, the mathematical expression for a sound wave without forcing is

$$\frac{1}{v^2}\frac{\partial^2 p(t,\vec{r})}{\partial^2 t^2} - \nabla^2 p(t,\vec{r}) = 0 \tag{2-2}$$

where v is the magnitude of the velocity vector.

2.2 Signal Processing

Nowadays, data management is something very important as we produce a lot of data per second, but it is not only a problem of today. Many centuries ago, mankind started to use signals to communicate faster and efficiently, so is not surprising that a mathematical theory emerged to solve communications problems.

This theory is the **information theory** or theory of communication, which has been worked and developed by different scientist such as Claude Shannon or Robert Craig, where the idea of signals have a relevant role.

The term signal is define as something that conveys information, any kind of information such as the state or behavior of a physical system, a more general definition of a signal is to think of it as a physical phenomenon that carries some information or data. The usual mathematical representation of a signal is a function, usually dependent on one or more independent variables, although time is often used, which depends on a variable that can be continuous or discrete. Continuous time signals are called *analog signals*: haven an infinite number of values, one value for all points in time at some (possibly infinite) interval, and are represented by a continuous time dependent variable. Discrete time signals usually called *digital signals*: have finite values, for only points in the discrete time domain, and are represented by a sequence of numbers. [**?**, **?**].

2.2.1 Time Domain

Audio

The discrete signal, as an audio signal, will be denote as $x[n]$, while the continuous signals as $x(t)$. The mathematical representation for a discrete signal can be written by either of the two following methods:

$$x = \{x_n\} = \{x_{-n}, x_{-n+1}, \cdots x_{-1}, x_0, x_1, \cdots, x_{n-1}, x_n\}$$
$$x_n = x[n] = x(nT), \qquad -\infty < n < \infty \tag{2-3}$$

or equivalent, moving the element 0 to 1 at $n = 0$

$$x[n] = \begin{cases} \left(\frac{1}{a}\right)^n & n \geq 0 \\ 0 & n < 0 \end{cases}$$

where $x_n = x[n]$ is the n^{th} number of the sequence, $x(nT)$ is the continuous analog signal, T is the *sampling period* and its reciprocal, $f_s = 1/T$ is the **sampling frequency**.

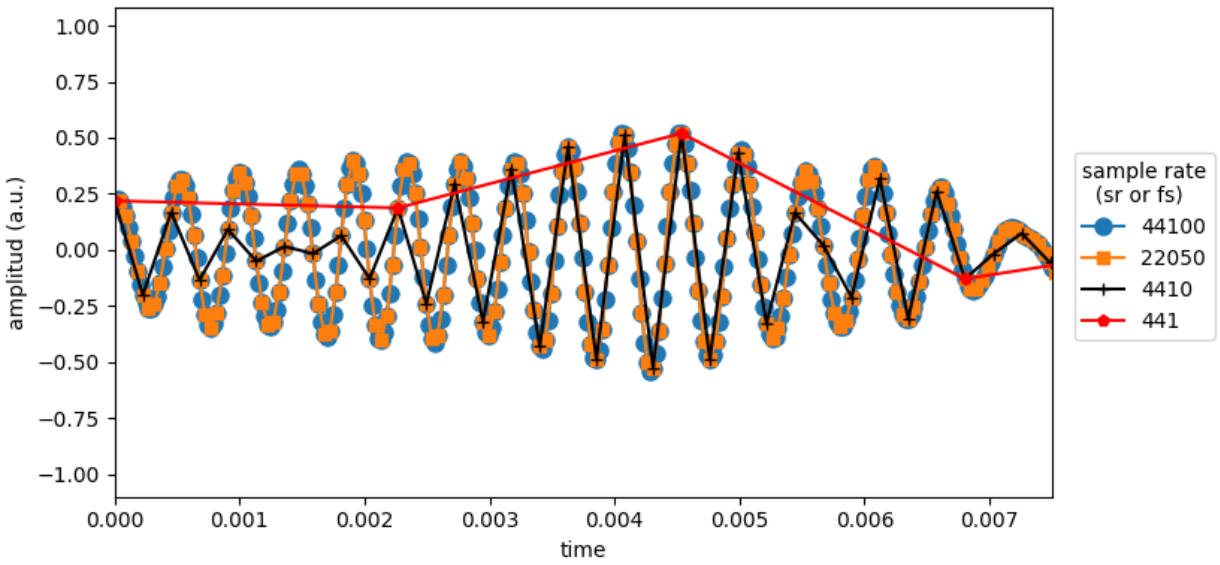


Figure 2-1: Sampling rate or sampling frequency of a continuous noisy signal. Note how with at lower sampling rates the sampled signal does not correctly reproduces the original signal. In order to do not to loss the function behavior, the sampling rate must satisfy the sampling theorem: the signal has to be sampled with at least twice the frequency of the original signal.

Envelope

The sound envelope describes how the amplitude of the sound changes over time. It is not an instantaneous measurement, it is a curve such that it covers the audio signal and is calculated

with its extreme points or percentiles. In this work, the envelope of a function $x(t)$ will be denote as $e(t)$.

2.2.2 Spectral Analysis

In the previous section sound was studied and defined as a sequence of numbers that usually depends of the time, this mathematical representation allows the visualization and modification of audio. This approach to study is called time domain analysis and exploring the audio by its amplitude, although sometimes is good enough in many other cases it is not enough.

A more powerful tool is the spectral analysis, where the signal of interest is studied in the frequency domain. This analysis aims to characterize the frequency content of a signal by decomposing it into an orthogonal basis of periodic functions: the object is to find the coefficients of the expansion that gives the amplitude (or weight) of its corresponding frequency.

Let us define some terms and spectral characteristic that this work used.

Fourier Series (FS)

The goal of the spectral analysis is to decompose the signal into its frequency, find the best basis and expansion coefficients for the signal. But, what is the best approximation?
The answer to this question will depend of the nature of the phenomena, the best basis is the function that best emulates the behavior of the phenomena. Therefore, it is natural to think that for audio signals, and many other digital signals, the best basis functions are trigonometric functions, which are periodic and easily differentiable.

Any function can be approximate by a linear combination of a basis function, in particular this functions can be a Laurent polynomial

$$f(x) \approx \sum_{n=-N}^{N} c_n q^n(x)$$

which are polynomials with positive and negative power terms. Although polynomials are solid basis, not all problems are well defined with them.

Other important basis functions are the trigonometric polynomials, these polynomials of degree N, $p : \mathbb{C} \to \mathbb{C}$, has the following form

$$p(x) = \sum_{n=-N}^{N} c_n e_n(x)$$

where $c_n \in \mathbb{C}$ are some coefficients to be calculated and $e_n(x) = e^{2\pi inx}$ is the basis function with $n \in \mathbb{Z}$. Note that the expansion basis is the exponential function which is related to the trigonometric function by the Euler's formula

$$\int_0^1 e^{2\pi inx} dx = \cos(2\pi nx) + i \sin(2\pi nx)$$

This base is a good election since it is orthonormal

$$e^{2\pi inx} \cdot e^{2\pi imx} = \delta_{m,n}$$

which implies

$$\int_0^1 p(x)\overline{e_n(x)}dx = c_n$$

It is very useful to calculating the coefficients, but f must be integrable if its domain is the complex set and a square-integrable function[2] if its domain is the real numbers. Then, the nth Fourier Coefficient of f is defined as

$$\hat{f}(n) = \int_0^1 p(x)\overline{e_n(x)}dx$$

In the same way, the Nth Fourier polynomial f_N of f is

$$f_N(x) = \sum_{n=-N}^{N} \hat{f}(n)e_n(x)$$

In other words, $f_N(x)$ is the the trigonometric polynomial of degree N whose coefficients are the Fourier coefficients $\hat{f}(n)$.

Using this polynomial, a "good approximation" means that f_N is an approximate function of f when $N \to \infty$, $\lim_{N\to\infty} f_N = f$.

$$f(x) = \lim_{N\to\infty} f_N(x) = \sum_{n=-\infty}^{\infty} \hat{f}(n)e_n(x) = \sum_{n\in\mathbb{Z}} \hat{f}(n)e_n(x) \tag{2-4}$$

although the equal symbol is present, it is an approximation of the function because the calculation of the series produces truncation errors. Using the Euler's relations

$$f(x) = \sum_{k=0}^{\infty} A_k \cos\left(\frac{2\pi k}{T}x\right) + \sum_{n=0}^{\infty} B_k \sin\left(\frac{2\pi k}{T}x\right)$$

where T is the common period of the function $f(t)$ and its the fundamental frequency is $f_1 = 1/T$. [**?**, **?**]

[2] a function is called square-integrable function if $\int_{-\infty}^{\infty} |f(x)|^2 dx < \infty$

Fourier Transform (FT)

This transformations is a generalization of the Fourier series (FS), it is the continuous analog; instead of calculating the discrete values of the Fourier coefficients $\hat{f}(n)$ with $n \in \mathbb{Z}$, the transformation takes continuous values $\hat{f}(x)$ for $x \in \mathbb{C}$. This transformation is defined as

$$\hat{f}(\omega) = \int_{-\infty}^{\infty} f(x) e^{-2\pi i \omega x} dx, \qquad f(x) = \int_{-\infty}^{\infty} \hat{f}(\omega) e^{2\pi i \omega x} d\omega \tag{2-5}$$

the right-hand side is the inverse transformation, this transformation recovers f from its Fourier coefficients $\hat{f}(\omega)$ with $\omega \in \mathbb{C}$.

The Fourier transformation is a linear transformation $\mathcal{F} : \mathbb{C} \to \mathbb{C}$of the function f, it will be denoted as $F(\omega) = \mathcal{F}\{f(x)\} = \hat{f}(\omega)$.

Discrete Fourier Transform (DFT)

The discrete Fourier Transform (FT) transform a sequence of N elements,$\{x_K\} := x_0, x_1, \ldots, x_{N-1}$, into another sequence of complex numbers, $\{X_K\} := X_0, X_1, \ldots, X_{N-1}$, which is defined as

$$X_k = \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi i}{N} kn} \tag{2-6}$$

$$X_k = \sum_{n=0}^{N-1} x_n \ \left[\cos\left(\frac{2\pi}{N} kn\right) - i \ \sin\left(\frac{2\pi}{N} kn\right) \right] \tag{2-7}$$

Fast Fourier Transform (FFT)

A fast and efficient algorithm to compute the discrete Fourier transform (DFT) or its inverse (IDFT). This algorithm rapidly computes such transformations by factorizing the DFT matrix into a product of sparse (mostly zero) factors. [**?**]

$$X_K = \sum_{n=0}^{N-1} x_n e^{-2\pi i k n / n}, \qquad k = 0, 1, \ldots, N-1 \tag{2-8}$$

where $x_k \in \mathbb{C}$ and $e^{-2\pi i/n}$ a primitive root.

Short-Time Fourier transform (STFT)

$$\textbf{STFT}\{x(t)\}(\tau, \omega) = \int_{-\infty}^{\infty} x(t) w(t-\tau) e^{-i\omega t} dt = \mathcal{F}\{x(t) w(t-\tau)\} \tag{2-9}$$

where $w(t)$ is the window function (commonly Gaussian, Hann, or Hamming windows)

$$\textbf{DSTFT}\{x(t)\}(n, \omega) = X(n, \omega) = \sum_{n=-\infty}^{\infty} x[n] w[n-m] e^{-i\omega n} \tag{2-10}$$

Convolution Theorem

This theorem states that under suitable conditions the Fourier transform of a convolution of two functions (or signals) is the pointwise product of their Fourier transforms.

$$\mathcal{F}\{f * g\} = \mathcal{F}\{f\}\mathcal{F}\{g\} = f * g \tag{2-11}$$

Spectrograms

A spectrogam is a graphic representation of a recorded sound (like a bird song) that includes information on both frequency and loudness, usually by plotting frequency against time and indicating intensity by shading or colors. It usually also called power spectrum and denoted by $S_{xx}(x(t))$ where $x(t)$ is the time series signal. Mathematically speaking, spectrograms are defined as the energy spectral density as follows

$$S_{xx} = spectrogram\{x(t)\}(\tau, \omega) = |X(\tau, \omega)|^2 \tag{2-12}$$

This visualizations are usually heatmaps, i.e., as an image with the intensity shown by varying the colour or brightness. When are plotted in three-dinemnsions are called waterfall plots.

Power Spectral Density (PSD)

It is a power spectrum (spectrogram) of a signal which describes the power present in the signal as a function of frequency, per unit frequency. Power spectral density is commonly expressed in watts per hertz (W/Hz) while power spectrum in dbS or dBFS.

Mel Spectrogram (MFCC)

The mel-spectrogram is often log-scaled before, it remaps the values PSD in hertz to the mel scale. MFCC is a very compressible representation, often using just 20 or 13 coefficients instead of 32-64 bands in Mel spectrogram. The MFCC is a bit more decorrelarated than other spectrograms representations which is beneficial for some application like Gaussian Mixture Models.

Fundamental Frequency

It is the lowest frequency produced by the oscillations of an object. The fundamental frequency is also called the first harmonic of the instrument or pitch.

Although there are many ways to compute it, the most used algorithm to the pithc extraction is the YIN (or the probabilistic version PYIN) algorithm. This algorithm compute an estimation of the fundamental frequency (F_0) of speech or musical by following the next steps:

Step 1: The autocorrelation method, uses the autocorrelation function (ACF).

Step 2: Difference function.

Step 3: Cumulative mean normalized difference function.

Step 4: Absolute threshold.

Step 5: Parabolic interpolation

Step 6: Best local estimate

This algorithm is already implemented and tested by librosa [**?**, **?**].

Harmonics

A harmonic is a wave or signal whose frequency is an integral (whole number) multiple of the fundamental frequency of the same reference signal or wave. The presence of harmonics in a signal depends of the nature of the signal. Birds and human vocalization usually have harmonics.

Mid Spacial Frequency (f_{msf})

The mid spacial frequency is defined as the average of the frequencies presented in the sound pondered by the sound energy.

$$f_{msf} = \frac{1}{E}\sum_{i=1}^{N} f_i \tag{2-13}$$

2.2.3 Acoustic Indexes

This indexes are used to characterize the signal properties, both in spectral and time domain

Power Spectral Entropy (SE)

$$PSE(F) = -\frac{1}{\log N_n} \tag{2-14}$$

Spectral Content Index (SCI)

Index with information about the spectral content of a signal. Simple signals has a SCI close to 1 while for complex signals the index value increase.

$$SCI = \frac{f_{msf}}{FF} \tag{2-15}$$

Acoustic Dissimilarity

Indexes used to compute how similar are two set of Fourier coefficients. Let x and y two N point Mel mean spectral of interest, $x, y \in \mathbb{R}^N$, then the acoustic dissimilarity can be calculated with any of the following indexes

- Correlation-based dissimilarity

$$correlation = \sqrt{1 - \frac{||x \cdot y||_1}{||x||_2 \; ||y||_2}} \tag{2-16}$$

- Symmetric Kullback–Leibler divergence

$$KL = \frac{1}{2} \sum_i^N \left(x_i \log_2 \left| \frac{x_i}{y_i} \right| + y_i \log_2 \left| \frac{y_i}{x_i} \right| \right) \tag{2-17}$$

- The integral of pointwise difference

$$D_f = \frac{1}{2} \sum_i^N |x_i - y_i| \tag{2-18}$$

Acoustic indexes definition taken from [**?**]

This indexes are from 0 to 1, where 1 represents great similar between Fourier coefficients.

2.3 Dynamical Systems

In general terms a dynamical system is any system, natural or man-made, that changes over time: the water cycle, the electron moving around proton, or even climate change... there are an endless numbers of examples and the birdsongs is one of them.

Formally speaking, a dynamical system is a system such that a set of variables uniquely define its state and with an associated **rule** to describe its behavior over time. **Differential equations** are used to study dynamical systems with physical modeling involved to describe the system at an instant in terms of some variables set called **state space**, or state variables, a n-dimensional point $x \in \mathbb{R}^n$ that depends of the complexity of the system. Depending on the nature of system, it can be described in discrete time steps or on a continuous timeline, but in either cases the state will be defined by a *difference equation* (or an iterative map)

$$\begin{aligned} x_t &= f(x_{t-1}, t) \qquad \text{(discrete)} \\ \frac{dx}{dt} &= f(x, t) \qquad \text{(continuous)} \end{aligned} \tag{2-19}$$

where f is a function defined by system evolution time rule. If the time of the system is continuous and deterministic, the system is defined by a **flow**, also called flow map and define as $x(t) = \phi_t(x(0))$. Dynamical system are either **deterministic**, given an initial point the final state is defined by a unique state, or **stochastic**, or random if there is a probability distribution. In any case, a dynamical system is described mathematically by an *initial value problem (IVP)* which implies a notion of time, in fact a state at one time will evolves to another possible state or collection of states, as a single quantity used to order the states chronologically. [**?**]

2.4 Bifurcation Theory

A bifurcation is the division of something into two parts. The same idea is used by bifurcation theory, which studies the mathematical changes in the topological or qualitative structure of a family of curves and the solutions of a family of differential equations, which studies how the sates of system bifurcate. In particular, for dynamic systems a bifurcation occurs when a small change in values of the system's parameters causes a sudden qualitative change in its behavior; for examples the dynamic system starts and ends to oscillating at certain critical points. [**?**]

A formal definition starts from considering a set of ordinary differential equations, $\dot{x} = f(x, \lambda)$, which depends of a n-dimensional variable $x \in \mathbb{R}^n$, p parameters $\lambda \in \mathbb{R}^p$, and a smooth time rule function $f : \mathbb{R}^{n+p} \to \mathbb{R}^n$. Then, a bifurcation occurs when λ takes a close value λ_1 such that the the number or stability of equilibria points or periodic orbits of f changes significantly.

An interesting way to think of bifurcations can be as a failure in the stability of the system within a family. To classify bifurcations we study their stability with the Kupka-Smale theorem which lists three generic properties of vector fields:

- Hyperbolic equilibrium points.
- Hyperbolic periodic orbits.
- Transversal intersections of stable and unstable manifolds of equilibrium points and periodic orbits.

In addition, the term **codimension** is used to describe the number of equality conditions that characterize the bifurcation, i.e. the minimum number of parameters if families in which the bifurcation take place.

A single failure in the Kupka-Smale properties yield to divide codimension one bifurcations in the following bifurcations types:

- Equilibria

  - Saddle - Node
  - Andronov-Hopf
- Periodic Orbits
  - Fold Limit Cycle
  - Flip Bifurcation (aka Period Doubling)
  - Neimark-Sacker Bifurcation (aka Torus)
- Global Bifurcations
  - Homoclinic Bifurcation of equilibria
  - Homoclinic tangencies of stable and unstable manifolds of periodic orbits
  - Heteroclinic Bifurcation of equilibria and periodic orbits

In this work, the dynamical system of study (the avian vocal organ) will present interesting behaviors with Saddle-Node, Hopf, and limit cycle bifurcations. The global bifurcation involving all these bifurcations is the Bogdanov–Takens bifurcation that will be explained in section 2.4.5.

Any study of bifurcation begins by examining the **equilibrium points** (also called equilibrium or *fixed points*) of the dynamical system. These points are generated by the system of ordinary differential equations (ODEs) and are time-invariant solution. Mathematically speaking, the ODE $x' = f(x)$ has an equilibrium solution (or a steady state) $x(t) = x_e$ if this point is a root of the ODE function $f(x_e) = 0$. Although the process is fairly straightforward, it is not always easy to solve the equation $f(x) = 0$ and it is only possible in some special cases. [**?**]

2.4.1 Types of Equilibria

To equilibria of the bifurcations are studied by calculating the egienvalues and eigenvectors of the time function rule f with the Jacobian matrix associated to the dynamical system, defined as

$$J(x) := \frac{\partial f_i}{\partial x_j} = \begin{pmatrix} \nabla^T f_1 \\ \vdots \\ \nabla^T f_n \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_n} \end{pmatrix} \tag{2-20}$$

where $\nabla^T f_i$ is the transpose of the gradient of the i component. By evaluating the Jacobian matrix at the equilibrium point, all derivatives, and then calculating its eigenvalues it is possible to determine the linear stability properties of the equilibrium point. Then, the Jacobian has same number of egienvalues as the dimension of the variables vector; if all egienvalues have negative parts the equilibrium is called **asymptotically stable**, if at least one eigenvalue has positive real part is called **unstable**, and it is said to be **hyperbolic** if all the eigenvalues have non-zero real

parts, or **non hyperbolic** if at least one eigenvalue is null or has a zero real part.

While hyperbolic equilibria are robust, small perturbations does not causes qualitative changes to the phase space near to the equilibria point, non hyperbolic equilibria are not, small perturbations can cause a local bifurcation that change stability, vanish or split into many equilibria points, and are called *saddle-node equilibrium*.

Let us consider some examples of spaces for usual bifurcations.

One-Dimensional Space

Let us consider a scalar dynamical system defined by a differentiable function $f \in C^1$, $x' = f(x)$, with $x \in \mathbb{R}$. The equilibria points are the root of the function f, as shows Figure 2-2 where the first two points are hyperbolic and the other non-hyperbolic[3]. If the Jacobian $J = f'(x)$ at the root is negative it is a stable point $f'(x) < 0$, while if it is positive at the point $f'(x) > 0$ it is an unstable point.

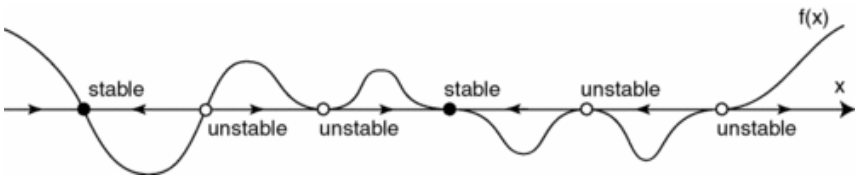


Figure 2-2: Diagram bifurcation for a one-dimensional dynamical system $x' f(x)$, the equilibrium points are the roots of f.

Two-Dimensional Space

A more interesting case is two-dimensional bifurcations. Let us consider a two-dimensional dynamical system defined as a set of ordinary differential equations

$$\begin{matrix} x_1' = f_1(x_1, x_2) \\ x_2' = f_2(x_1, x_2) \end{matrix}, \qquad J(x_1, x_2) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} \end{pmatrix}$$

Since the space is two-dimensional the Jacobian matrix has two eigenvalues that are either both real or complex-conjugate, defined by the trace and determinant of the Jacobian defined as follows

$$\tau = \mathrm{tr} J = \frac{\partial f_1}{\partial x_1} + \frac{\partial f_2}{\partial x_2}$$
$$\delta = \det J = \frac{\partial f_1}{\partial x_1}\frac{\partial f_2}{\partial x_2} - \frac{\partial f_2}{\partial x_1}\frac{\partial f_1}{\partial x_2}$$

[3]because the y-component of the slope is zero, the eigenvalue of the Jacobian matrix

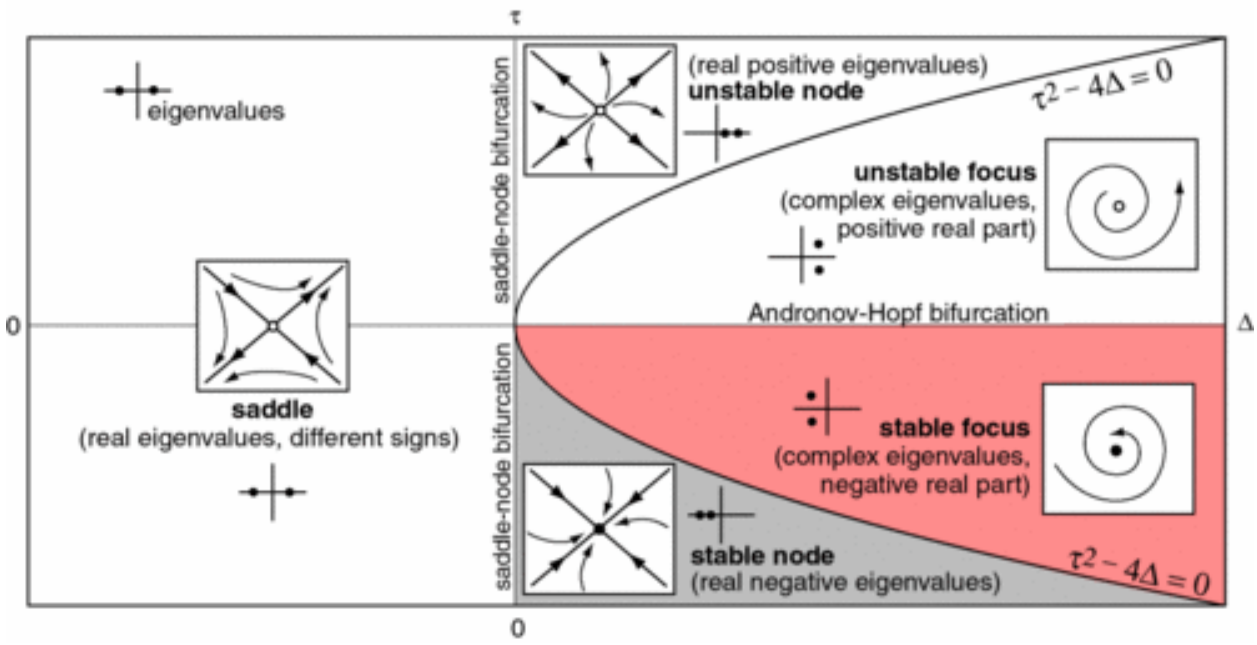


Figure 2-3: Two-dimensional dynamical system types of equilibrium points that depends of the Jacobian trace ($texttr$) and determinant (Δ). The red and gray shadow regions corresponds to stable equilibrium points.

Both hyperbolic and non-hyperbolic equilibrium points are present in this dynamical system: nono-hyperbolic equilibria are located on the half-axis $\tau = 0$, $\Delta > 0$, and on the axis $\Delta = 0$ arising at the Andronov-Hopf and at Saddle-Node bifurcation, respectively; while a hyperbolic equilibrium can be:

- **Saddle**, both eigenvalues are real and of opposite signs then the point is unstable.
- **Node**, both eigenvalues are real and of positive signs. The point is stable is the eigenvalues are negative and unstable when they are positive.
- **Spiral point (or focus)**, the eigenvalues are complex-conjugate. It is stable if the eigenvalues have negative part and unstable when they have positive part.

Figure 2-3 summarize the types of equilibrium for a two-dimensional dynamical system.

Three-Dimensional Space

In this case the Jacobian matrix has three eigenvalues, one of them must be real and the other two can be either both real or complex-conjugate. Figure 2-4 shows the interesting possible cases for this dynamical system that are defined by the type and sign of the eigenvalues. Similar to two-dimensional dynamical system, a hyperbolic equilibrium can be:

- **Node**, all eigenvalues are real and have the same sign. If the eigenvalues are positive (negative) the point is unstable (stable).

- **Focus-Node**, all eigenvalues have real parts of the same same where one eigenvalue is real and the others are a complex-conjugate pair. It is a stable (unstable) point when the sign is negative (positive).

- **Saddle**, all eigenvalues are real and at least one of them is positive and at least one is negative. They are always unstable points.

- **Saddle-Focus**, one real eigenvalue with the sign opposite to the sign of the real part of a pair of complex-conjugate eigenvalues. It is always an unstable point.

If time is reversed, changing $t \to -t$, the node and focus-nodes change their stability, while in contrast saddle and saddle-focus remain unstable.

Non-Hyperbolic Equilibria

These are not the only types of non-hyperbolic equilibria, those points having at least one eigenvalue with zero real part, three other examples in $\mathbb{R}^2$ are displayed in the Figure 2-5.

- **Center equilibrium**, a pair of purely imaginary eigenvalues. These points in linear system have families of concentric periodic orbits.

- **saddle-node equilibrium**, one zero eigenvalue in a nonlinear system undergoing a saddle-node bifurcation, the equilibria is always unstable. Here a saddle and node approach each other and merge into a single equilibrium point and then disappear.

- **Bogdanov-Takens equilibrium**, two zero eigenvalues of a nonlinear system that usually undergoes a Bogdanov-Takens bifurcation. As well as saddle node it is an unstable equilibrium.

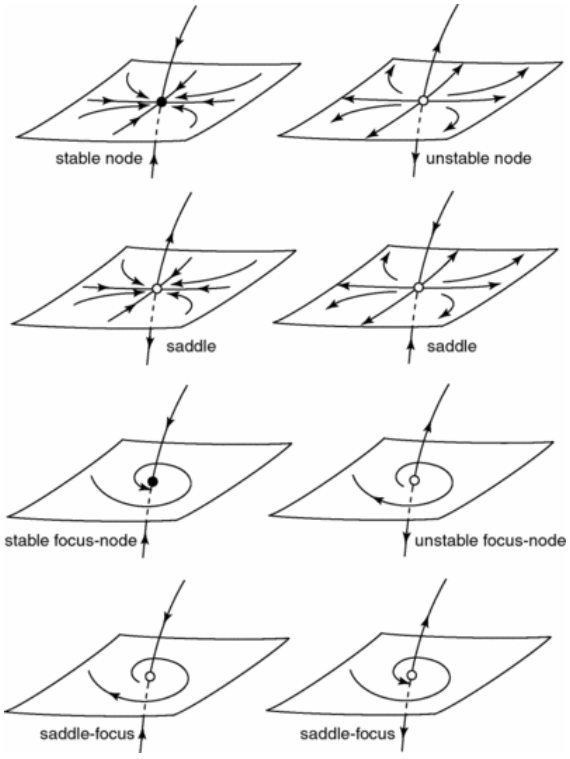


Figure 2-4: Equilibrium points examples for a three-dimensional dynamical system.

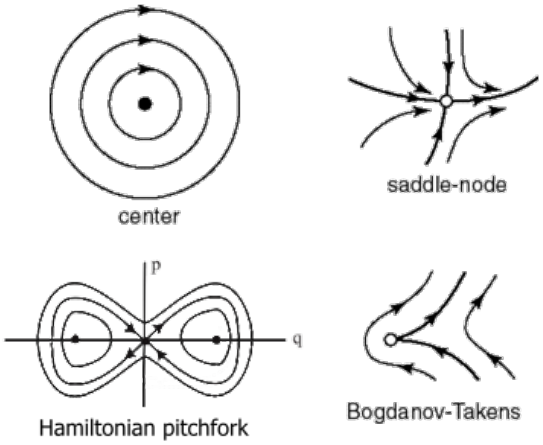
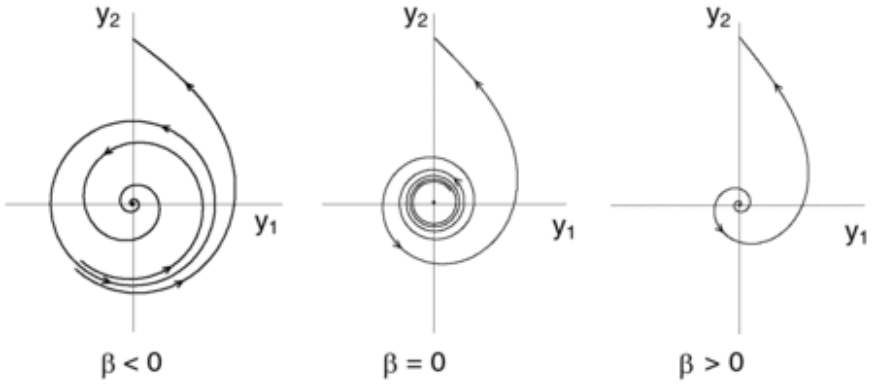


Figure 2-5: Non-hyperbolic equilibrium points in $\mathbb{R}^2$. The center and Bogdanov-Takens bifurcation diagrams correspond to Hamiltonian equations, $x'' = -\omega^2 x$ and $x'' = Kx^2$, respectively.

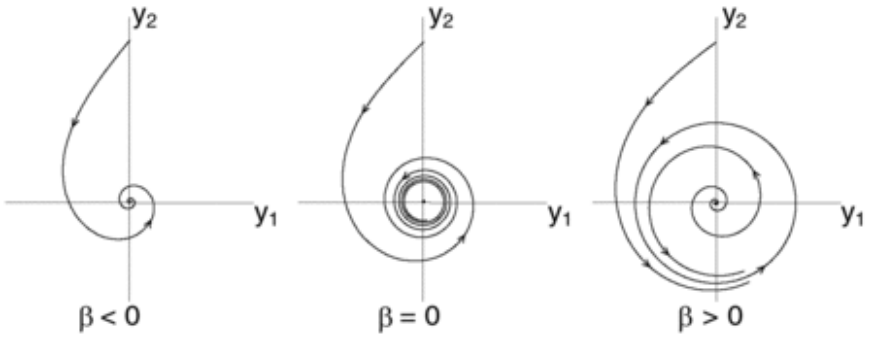
2.4.2 Andronov-Hopf

This bifurcation consist of a change in stability by a pair of pure imaginary eigenvalues that generate a limit cycle solution[4]. There are two possible types of limit cycles: **supercritical** and **subcritical**, stable and unstable respectively, Figure 2-6.

4a special type of solution for a dynamical system that repeats itself in time



(a) Supercritical bifurcation



(b) Supercritical bifurcation

Figure 2-6: Hopf possible types of bifurcations.

2.4.3 Saddle-Node

This bifurcation occurs when the critical equilibrium has one zero eigenvalue. It is also called fold or limit point bifurcation and consists in the creation, collision, and destruction of two critical points, as shows the Figure 2-7. .

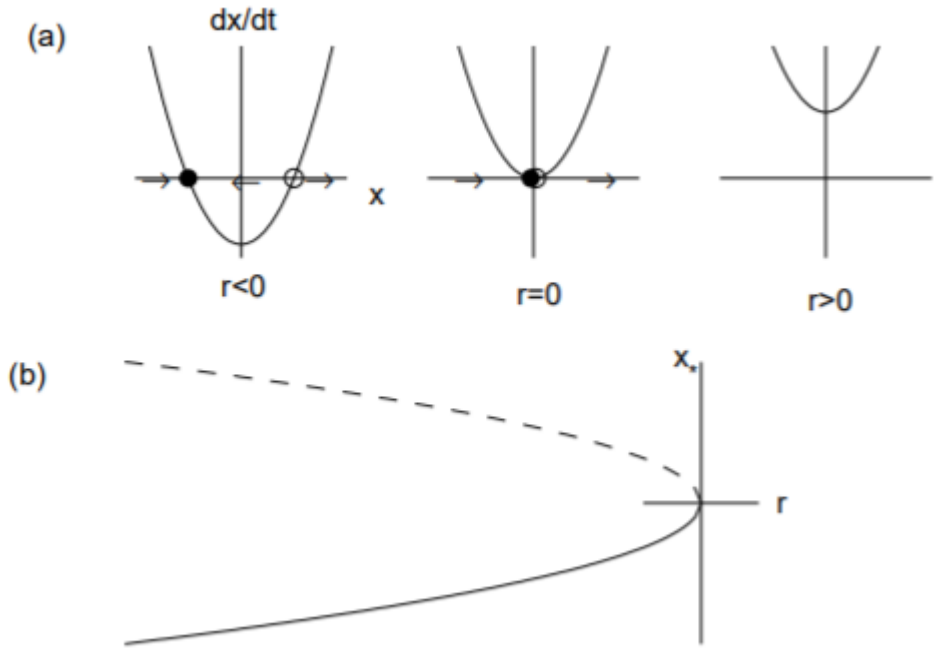


Figure 2-7: One dimensional saddle-node bifurcation $x' = r + x^2$. (a) Phase space, x' vs x. (b) bifurcation diagram

2.4.4 Limit Cycle

When the solution of the dynamical system is a stable periodic orbit, it is called a limit cycle or oscillator. This bifurcation is fascinating as they fill generate oscillating dynamical systems.

The stability of a periodic orbit is calculated using Poincare maps. If all eigenvalues have a modulus value less than unit the corresponding bifurcation is asymptotically stable, but if instead its modulus is greater than unit then the bifurcation is unstable.

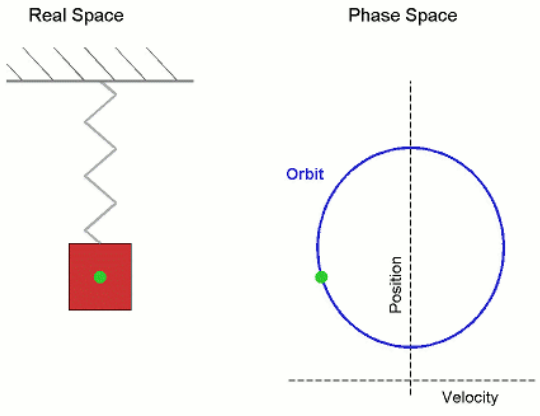


Figure 2-8: Periodic orbit shown in phase space, simple harmonic motion

2.4.5 Bogdanov–Takens (BT) Bifurcation

It consists of an equilibrium point that bifurcates into a two parameter family solutions of the ODE (2.4.5). The equilibrium point has two degenerated zero eigenvalues, of multiplicity two.

The ordinary differential equation system is

$$\begin{aligned} x_1' &= x_2 \\ x_2' &= \alpha + \beta x_1 + x_1^2 - x_1 x_2 \\ (\text{or } x_2' &= -\alpha - \beta x_1 + x_1^2 - x_2^3 - x_1 x_2 - x_1^2 x_2) \end{aligned} \tag{2-21}$$

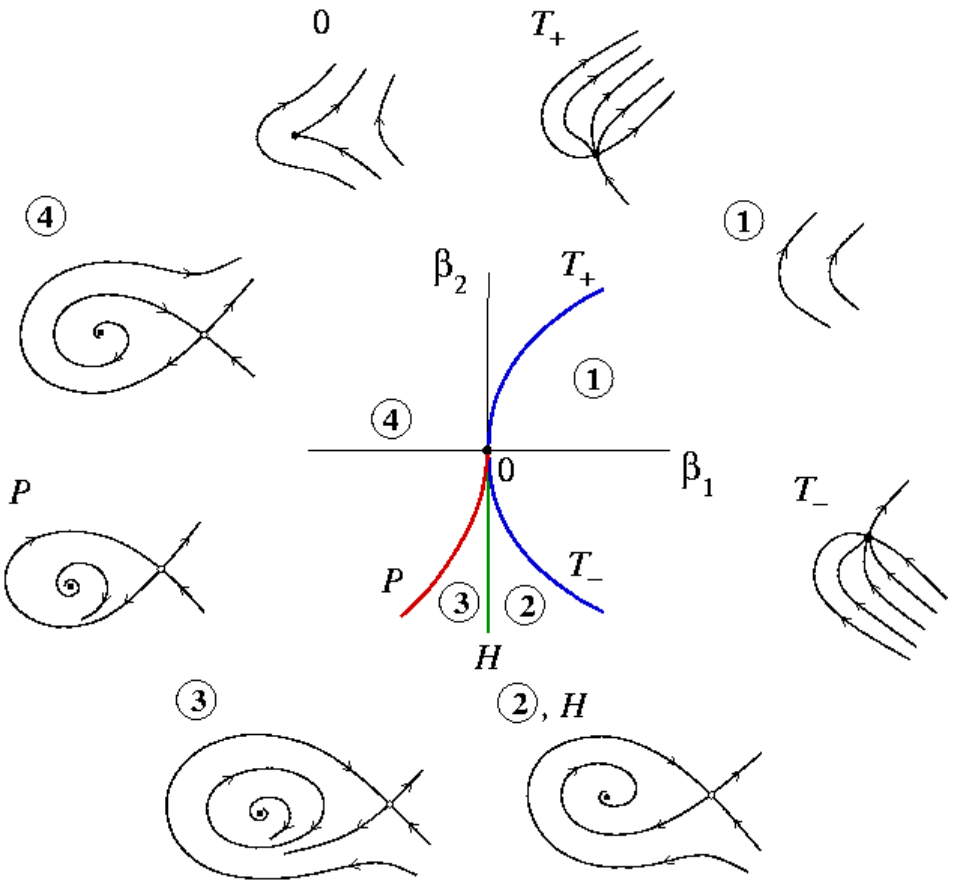


Figure 2-9: Bogdanov-Takens bifurcation, possible solutions of parameters families.

Close parameters values generate two equilibrium points, a saddle and nonsaddle, which collide and vanish via a saddle-node bifurcation. An Andronov-Hopf bifurcation is generated by the nonsaddle equilibrium that produces a limit cycle. This cycle degenerates into an homoclinic orbit to the saddle and disappears via a saddle homoclinic bifurcation. [**?**, **?**]

2.5 Numerical Optimization

Throughout the entire history of the universe nature and livings beings have been evolving to survive, to continue living as mush as possible, leading them to reach the current evolution

where we have develop wireless communications, for faster and easier interactions, flying machines, to travel huge distances in a few hours, and even bionic eyes, to see the quantum and cosmic world, but all these inventions have the same idea behind: *how to improve a something*: how to get betters results from a problem by understanding its nature and what it depends o?n. But, what does better results mean?

To define and quantify the phrase "better results", the theory of mathematical optimization emerges and creates a robust mathematical tool, with many concepts, algorithms, and theorems, to solve any kind of optimization problem, or any problem that can be written as one.

This chapter is an introduction to numerical optimization, defining and discussing the necessary concepts and theory used in this work. [**?**]

2.5.1 Introduction

First, any optimization problem must have a defined **objective function**: a quantitative score that measures the performance variable of the system of interest, usually a profit, time, energy, or any physical variable of the system; it depends on system characteristic which are called **variables**. The goal of optimization is to find **optimal** variables that minimize (or maximize) the objective function subject to some *constraints* that the system must satisfy.

In many cases the domain of the problem variables are constraint to some region, the variables limits are well determined and it is possible to define the **feasible domain** on the problem. The constraints and objective function can be linear, quadratic, or nonlinear functions, and may depend of some known values **parameters** which characterize the system under study.

In optimization, **modeling** is the process of identify the objective function, constraints and system variables. Any optimization problem has behind it a modeling process, the most important step, where the limits and scopes of the model are determined. Once the problem has been modeled, the next step is to apply an optimization algorithm to find the **optimal variables** and check if they are indeed the solution of the problem by means of some mathematical expressions called **optimal conditions**. At this point the problem has an optimal set of variables but sometimes it is not enough. To improve the modeling process a **sensitive analyze** can be used, which reveals how sensitive is the model to changing input data or parameters values.

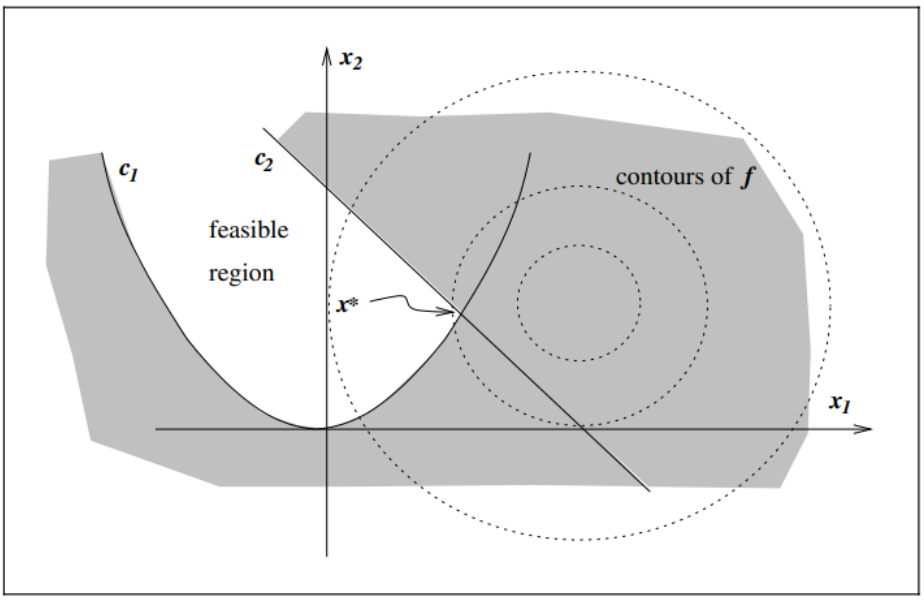


Figure 2-10: Geometrical representation of an optimization problem. [**?**]

Therefore, the optimization objective can be summarized as the minimization or maximization of a function whose variables must satisfy some constraints. Although the objective function have to be scalar, many problems functions can be written as vector functions and then the problem can though of as minimizing (or maximizing) a vector norm. Depending of the nature an behavior of the objective function and constraints, whether they are linear, quadratic, non smooth, or nonlinear functions, the optimization algorithm is chosen. With these properties, optimization theory classifies and studies the problems, in addition, the presence (or not) of constraints is also important and allows another classification to the problems, *constraint or unconstrained optimization*.

Mathematical Formulation

Using mathematical notation, an optimization problem can be written as follows

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x) \\ \text{subject to} \quad & c_i(x) = 0, \quad i \in \mathcal{E}, \\ & c_i(x) \geq 0, \quad i \in \mathcal{I} \end{aligned} \tag{2-22}$$

here x is the vector of variables (also called unknowns or parameters), $f : \mathbb{R}^n \to \mathbb{R}$ the scalar objective function, c_i scalar constraints functions that that must satisfies the vector variables where $\mathcal{E}$ and $\mathcal{I}$ are the indices for equality and inequality constraints, respectively.

Convexity

An important property of a function or set in optimization is the *convexity*. In fact, a wide variety of problems posses this property that make them easier to solve in theory and in practice, they are studied and solved with the theory of **convex optimization**.

A set S is called convex if the straight line segment connecting any two points of S lies entirely inside S, formally speaking for any two points $x, y \in S$, then $\tau x + (1-\tau)y \in S$ for all $\tau in[0,1]$. Similarly, a function f is convex if its domain S is a convex set and if for any two point of $x, y \in S$, the following property is satisfied

$$f(\tau x + (1-\tau)) \leq \tau f(x) + (1-\tau)f(y), \quad \text{for all } \tau \in [0,1] \tag{2-23}$$

This property is important because if the objective function and its domain satisfy it, then any local solution of the problem is in fact a global solution. For the special case where everything is convex, the objective function and constraints, the term *convex programming* is used.

Optimal Solution

Depending of the nature and properties of the function, the problem would have a **global minimizer** (**maximizer**) or **local minimizer** (**maximizer**), the optimal solution. The ideal case is to find a global optimizer of the objective function f. It is define as point x^* such that

$$f(x^*) \leq f(x) \forall x \in D \quad \text{(global minimizer)}$$
$$f(x^*) \geq f(x) \forall x \in D \quad \text{(global maximizer)}$$

where D is the problem domain of interest which can be the a real set $\mathbb{R}$. Sometimes it is difficult to find the global optimizer since our knowledge of the objective function is only local. Usually a local optimizer is sufficient for some problems and the most optimization algorithms are able to find them. A point x^* is called **local optimizer** (minimizer or maximizer) if there is some neighborhood N such that

$$f(x^*) \leq f(x),\ \forall x \in N \quad \text{(local minimizer)}$$
$$f(x^*) \geq f(x),\ \forall x \in N \quad \text{(local maximizer)}$$

where this neighborhood is an open set that contains x^*. When the point satisfied any of this inequalities is called *weak local optimizer*. Similarly, a **stric local optimizer**, the best optimal of the neighborhood, is a point such that

$$f(x^*) < f(x) \forall \in N, \text{ with } x \neq x^* \quad \textbf{(stric local minimizer)}$$
$$f(x^*) > f(x) \forall \in N, \text{ with } x \neq x^* \quad \textbf{(stric local maximizer)}$$

Therefore, to find an optimal point a basic but expensive algorithm can be scan the whole feasible region, dividing this region into a grid of points and computing the objective function at each point. This technique for solving the problem is a **brute-force search**, it is easy to implement and always finds a solution if one exist, but the implementation is very expensive as it depends of the number of candidates.

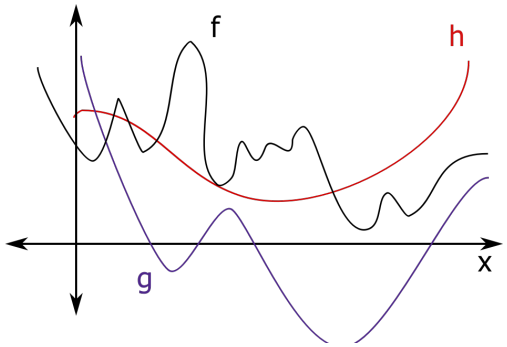


Figure 2-11: Simple and difficult objective functions to calculate a global minimizer

When the objective function is **smooth** or **convex**, the optimal can be found in a better way. In particular, if the function is twice continuously differentiable, the gradients (∇f) and Hessian ($\nabla^2 f$) of the function will tell us whether x^* is a local optimizer by checking some conditions. Another mathematical tool used to study smooth optimization is the Taylor's theorem. [**?**]

2.5.2 Inverse Problem

A good physical theory will not only describe the phenomena but also be able to make predictions about some variables of the system, if the problem is completely described by the physical model we can predict the results for some measurements of the system. This problem procedure is called a **forward problem**, it calculates the result of the model with some measured samples. While on the contrary, an **inverse problem** consists of using the outcomes of a physical model, the result, to infer the values of the parameters that characterize the physical system. The major difference between forward and inverse problem is uniqueness of the solution, while the forward problem has a unique solution, since forward problems use deterministic physical models, the inverse problem does not. [**?**]

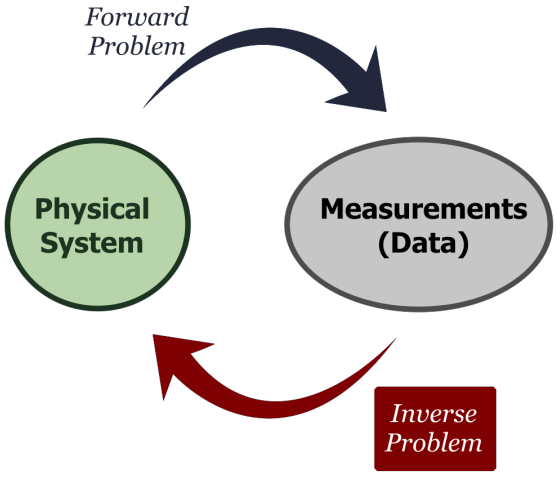


Figure 2-12: Diagrams of inverse and forward problem.

The notion of an inverse problem has a probabilistic nature, they are problems that involve mea-

surements, and so a good approach is to study them on the basis of conditional probabilities and Bayes's theorem. Another simple way to define an inverse problem is to use the probability theory: given some measurements we are interested in to find the distribution that characterizes them.

Note: An inverse problem can be formulated as a minimization problem
Let us consider the matrix equation problem $Ax = b$, with $A \in \mathbb{R}^{m \times n}$ a matrix transformation, $x \in \mathbb{R}^n$ the system variables vector, and $b \in \mathbb{R}^m$ the vector with known measured data. This is an inverse problem since we are interested in finding the system variables that best reproduce the known measured data taking into count the matrix rule transformation.

The naive solution to this equation is $x = A^{-1}b$, if A is well conditioned and then has an inverse, or using the Moore–Penrose inverse another solution is $x = (A^TA)^{-1}A^Tb = A^+b$. In either case the solution of this problem is to find the optimal vector x, that best reproduces the measurements, making use of the matrix rule transformation A and the known data b.

The minimal distance is obtained when b is projected onto the column space of A

$$b^* = Ax^*$$
$$dx^* = Ax^* - b^*$$

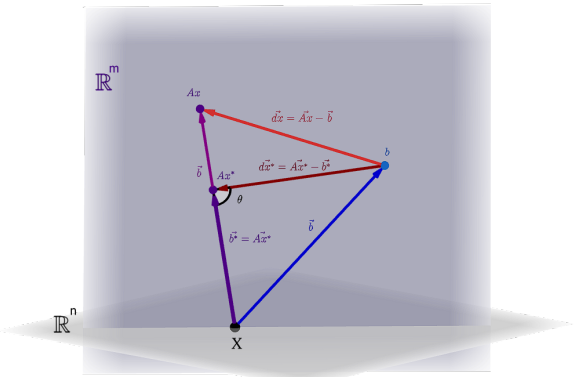


Figure 2-13: Linear transformation

where b is orthogonal to the projection b^* and Ax^*. Therefore, x^* is a solution if and only if dx^* is orthogonal to the column space of A

$$A^T dx = 0$$
$$A^T(Ax - b) = A^TAx - A^Tb = 0$$
$$A^TAx = A^Tb \quad \Rightarrow \quad A = (A^TA)^{-1}A^Tb$$

where again the solution is the Penrose inverse times the vector b which is applicable to all least squares problems for all possibles n and m.

This example shows how a linear transformation problem, an inverse problem, can be formulated as an optimization problem, minimizing the distance between two vectors, which may or may not be subject to some constraints.

$$Ax = b \qquad \leftrightarrow \qquad \underset{x \in \mathbb{R}^n}{\text{argmin}} \quad ||Ax - b||_2^2$$

2.5.3 Non-Linear Optimization

The optimization literature has shown how several phenomena can be modeled with linear programming, with some good enough tolerance, but in other situations it is not sufficient. When the problem is complex to model and has a complicate objective function and constraints, linear programming is deficient and a new tool must be used, **nonlinear programming**. *General nonlinear optimization* studies problems which both the objective function and functional constraints may be nonlinear functions, although the variables may have simple bounds associated. A more formal definition of a non linear problem is the following:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x) \\ \text{subject to} \quad & c_i(x) = 0, \quad i \in \mathcal{E}, \\ & c_j(x) \geq 0, \quad j \in \mathcal{I} \\ & l \leq x \leq u \end{aligned} \tag{2-24}$$

where again $f : \mathbb{R}^n \to \mathbb{R}$, ⌉ and $\mathcal{I}$ are the set of indices for equality and inequality constraints, respectively. The constraints are vector functions, $c_i : \mathbb{R}^n \to \mathbb{R}^m$ and $c_j : \mathbb{R}^n \to \mathbb{R}^p$ with $m = card(\mathcal{E})$ and $p = card(\mathcal{I})$, respectively. The lower bounds l and upper bounds u satisfies the natural conditions $-\infty < l_i \leq u_i < +\infty$ for all $i = 1, \ldots, n$, both bound are vectors $l, u \in \mathbb{R}^n$.

To guarantee solutions there are two types of optimality conditions: *necessary optimality conditions* which any solution point must satisfy, and the *sufficient optimality conditions* where if they are satisfied for a given point they guarantee that this point is in fact a solution. The necessary conditions are also called *first order*, since they involve the gradient of the objective function and constraints. The **KKT** (Karush-Kuhn-Tucker) conditions are the cornerstones of many nonlinear optimization algorithms. Second, the *second order conditions*, both necessary and sufficient, will study the second derivatives of the objective function to deduce useful information for finding the optimal point.[**?**]

3 The Problem and its Background

3.1 Birdsong

Just like humans, birds made vocalization to communicate each other using the **syrinx** to produce sound instead of mammals vocal folds. These vocalizations are called birdsongs and are used at most by bird males for attraction mate or territorial defense, females also generate birdsongs but shorter and simpler than males birdsongs. Depending of the bird specie the birdsong is learned from a tutor or just inherit to them, they born with the ability to sing, but always is a signature of the bird, a way to define and characterize the individual bird.

3.1.1 Bird Calls and Syllables

Not all vocalizations are the same. Some of them are shorter and sharper and less rhythmic, they are called **bird calls** and are used to communicate with other birds of the same or different species, to be in contact with their families and protect them.

On the other hand, birdsongs are usually generated by male and are more musical and longer than bird calls. In order to study them, there are three important features to be consider: sound quality, pitch trend, and number of sections.

To describe the sound quality of a birdsong the terms used are:

- **Clear:** Where the pitch is well defined, this is something you could sing.
- **Buzzy:** Something like a bee, the pitch looks distorted as if it have noise.
- **Thrilled:** Many syllables in a sound that are too fast to count (11 elemental sounds per second).

Some birds sing with more than one quality. As an example some sings with a part buzzy and another one clear, this type is called **Partly Buzzy** .

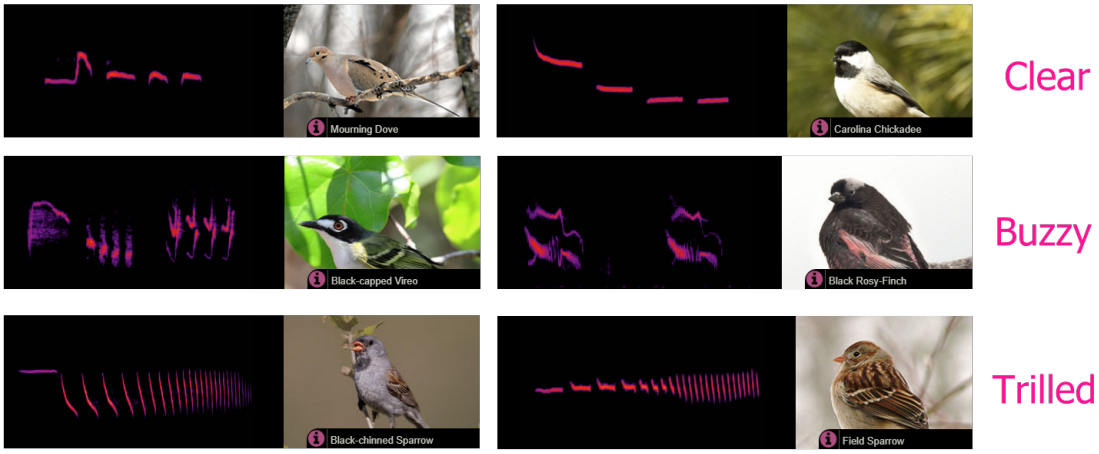


Figure 3-1: Classification by song quality.

What we call the signature of the bird is the pitch of the spectrum, even if each bird has a different spectrum they are usually classify by its overall behavior: is it steady, rising, falling, or varying, moving up and down?

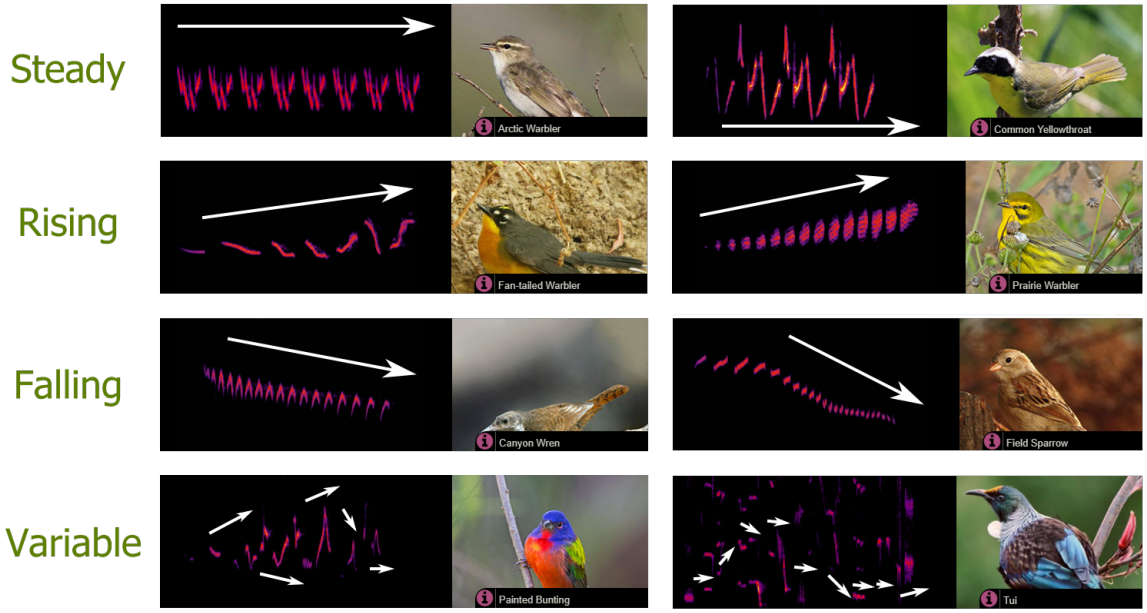


Figure 3-2: Classification by pitch trends.

In addition, the birdsongs can be break down into parts called sections, intervals where the pitch behaves in the same way, they are defined by the drastically changes in the pitch and are very utile to identify the bird. These section can be also break down into **syllables** (or elements) and **phrases**, group of consecutive syllables.

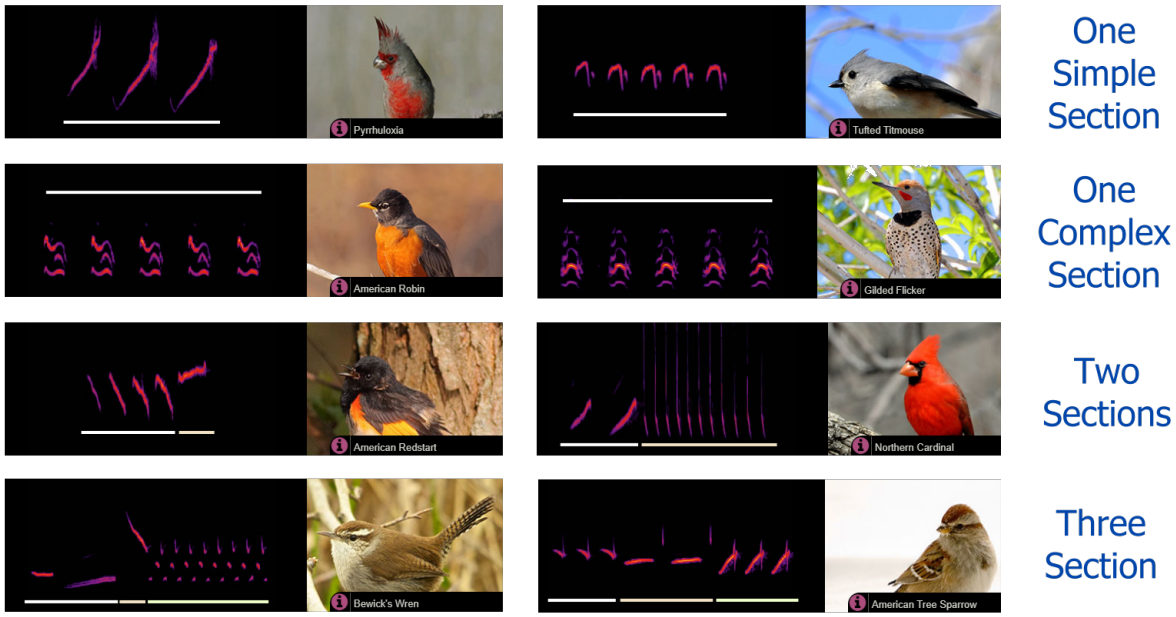


Figure 3-3: Classification by number of section. This classification can be done even when the song spectrum is complex, where the signature is not quite simple and may be a composition of two pitches since many birds have two independent syrinx organs.

All birds spectrograms was taking and adapted from the song learning game Bird Song Hero [**?**]. To deeper syllables study check the blog EARBIRDING, there you will find information about syllables classification with audio and spectrum examples and also a demonstration of how some birds of the same specie has similar spectrums.

3.2 Sound Production of a Birdsong

To understand how birds sing first let us study their anatomy, in particular the respiratory system.

Despite the fact that the respiratory system has many actors, to produce sound not all of them contribute. The organs involved to generate sound are the air sacs, bronchi, trachea, beak, and the **syrinx**, the main character in the sound production process.

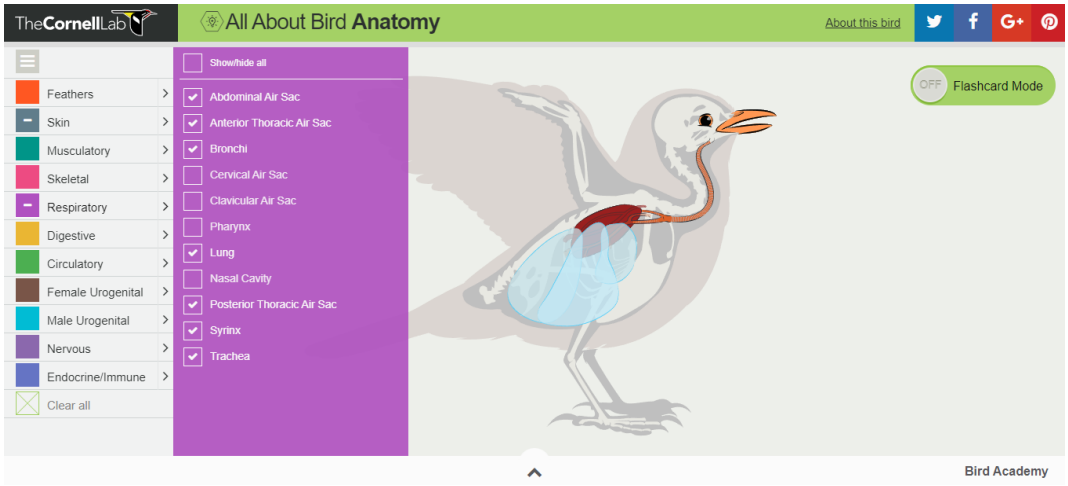


Figure 3-4: Bird respiratory system anatomy. [?]

3.2.1 Syrinx

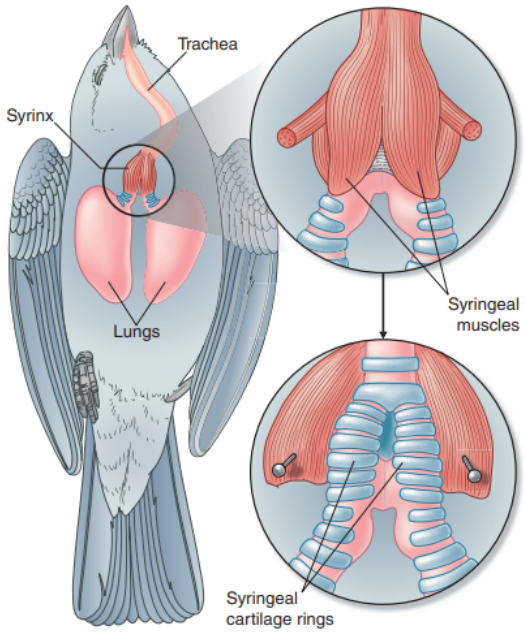


Figure 3-5: Sound production actors: **the syrinx**, air sacs, trachea, and beak. Image taken from [?].

The syrinx is primary sound-producing organ that birds use to generate birdsongs; it is equivalent to the mammals voice box, larynx, and unlike humans is located where the trachea forks into the lungs, the junction where the trachea (windpipe) splits to form the two tubular bronchi as that lead to the lungs as is showed in the Figure 3-5.

This organ consist of two independent smalls parts of tissue, known as **labia** or membrana tympaniformis, located in major of cases at the end of the bronchus. Each labia has six pairs of muscles, called syringial muscles Figure 3-6, that are controlled independently by a nervous system which allows to them make extraordinary and complex birdsongs, in fact they are great vocal gymnast.

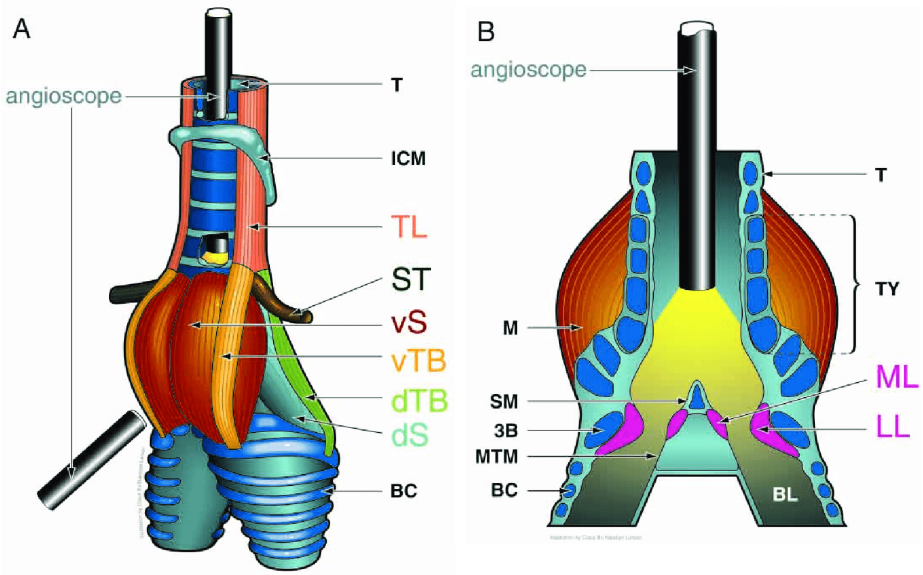


Figure 3-6: Syringeal muscles [**?**].

Since each labia is independently controlled, each one can produces different sounds making the bird singing different syllables at the same time, as shown in the Figures 3-8 and 3-7.

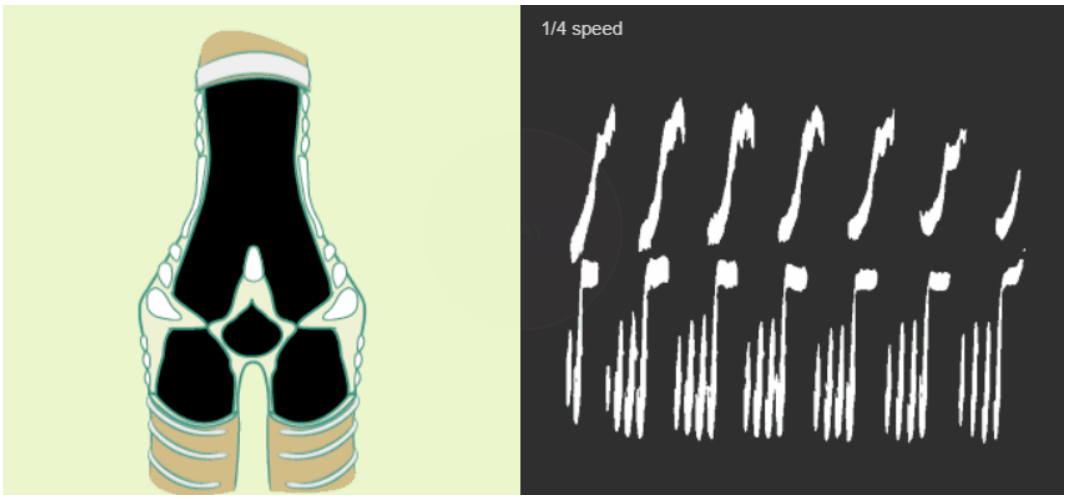


Figure 3-7: Independent syrinx oscillations and its corresponding spectrograms [**?**]

To produce the sound, the labia oscillations modulates the air passing through them, which comes from the air sac, and then goes to the trachea and expel by the beak, they attenuate the sound amplitude of a particular frequency and enhance the pitch.

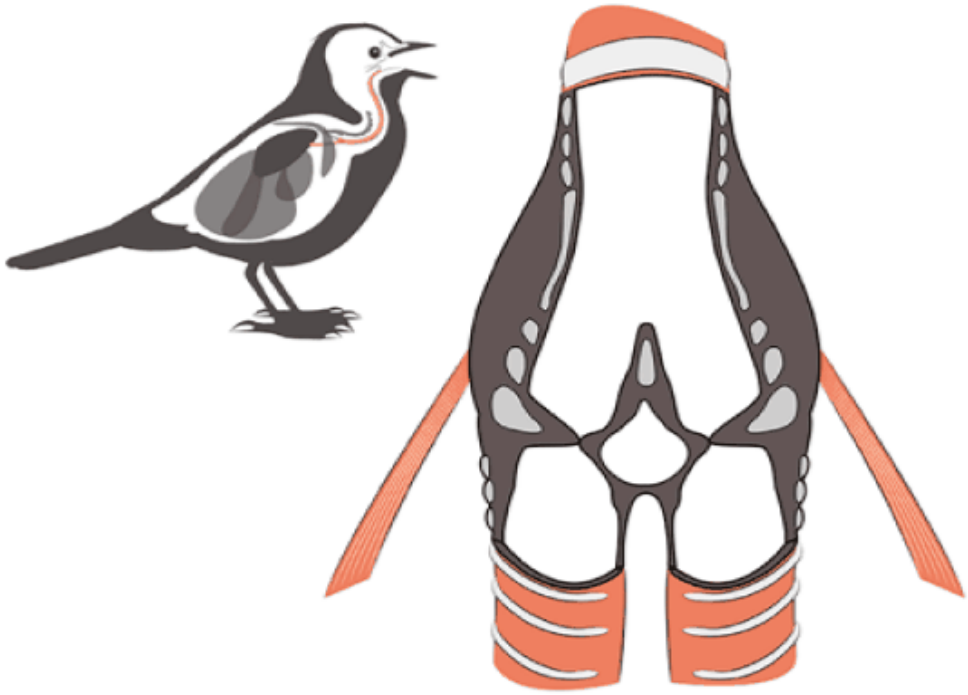


Figure 3-8: Syrinx oscillations. Many birds have two independent syrinx to modulates the air and produces birdsongs, some of them are able to control each syrinx an generate complex sounds as is showed in this figure.

3.2.2 Physics of a Birdsong

As was discussed in the section 2.1.1, a sound wave is the propagation of a pressure perturbation in an elastic medium such as air. This perturbation is produced by the vibration of molecules that generates local changes of density such that lead to changes in pressure, as is shown in the following Figure **3-9**. The density perturbation is proportional to the air displacement velocity and the initial density.

The pressure perturbation is generated by a vibrating sound source that moves forward and back to spread out the first medium particles, then they push their surrounding particles and the perturbation is propagated. The medium is a series of particles interconnected and interacting that must be able to transmit energy (an elastic medium).

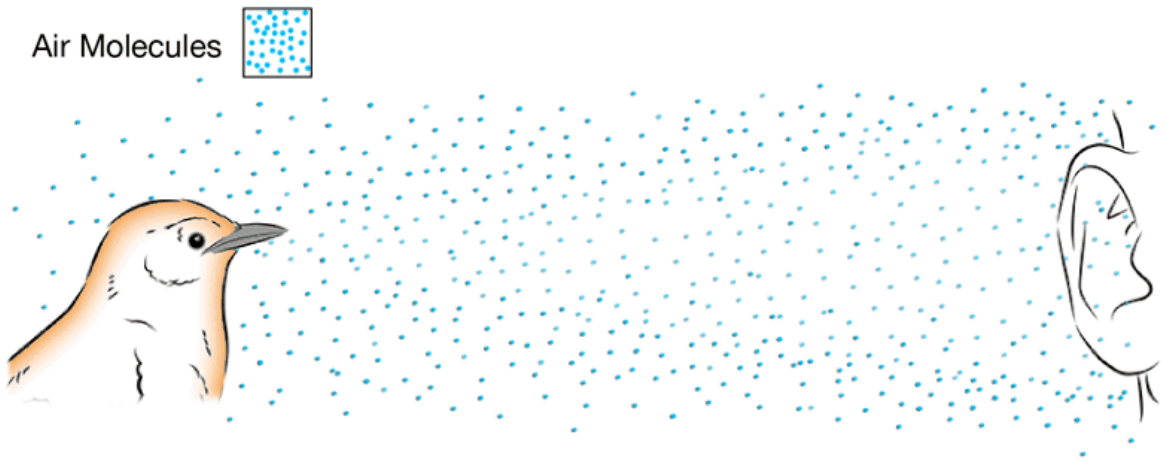


Figure 3-9: [?]

The usual sound wave representation is the **waveform** where the amplitude of the wave is plotted as a function of time, the amplitude represents the sound loudness. Another important sound feature is the frequency, that have information about the pitch, that is define as the rate at which sound waves are produced, sounds with higher frequencies than 20.000 Hz are called ultrasonic while frequencies below are known as infrasonic. The frequency inverse is the period $T = 1/f$, the amount of time it takes to complete one wave cycle[1]

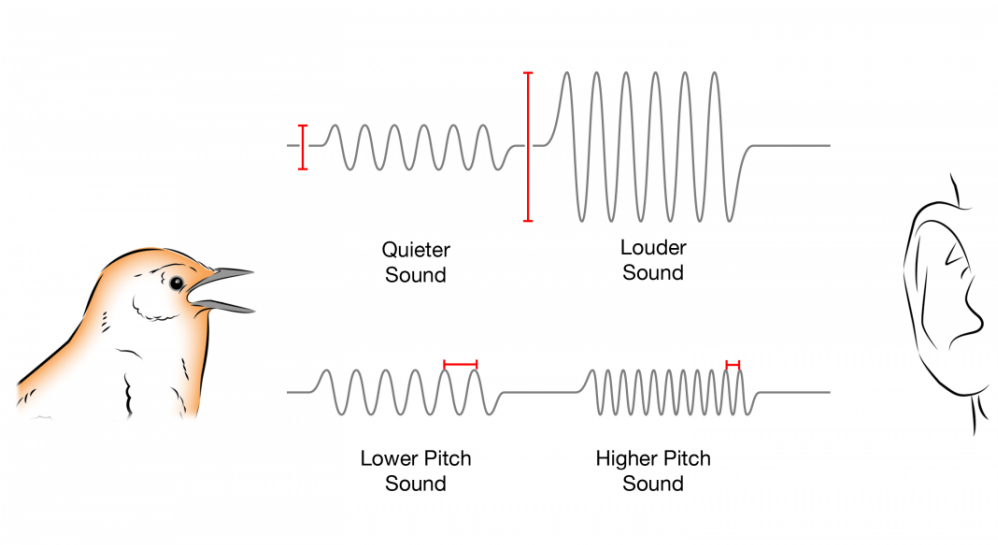


Figure 3-10: Sound pitch [?]

In physiology, sound is defined as how humans receive the sound waves and how is their perception by the brain. The human eardrums are night and day hit by sounds, as response they

[1]when two successive crests (or through) past an specific point

vibrate and convert these vibrations into electrical signals that travel to the brain and are interpreted as sound. The brain perception allows to us to distinguish and classify the signal by their pitch or loudness.

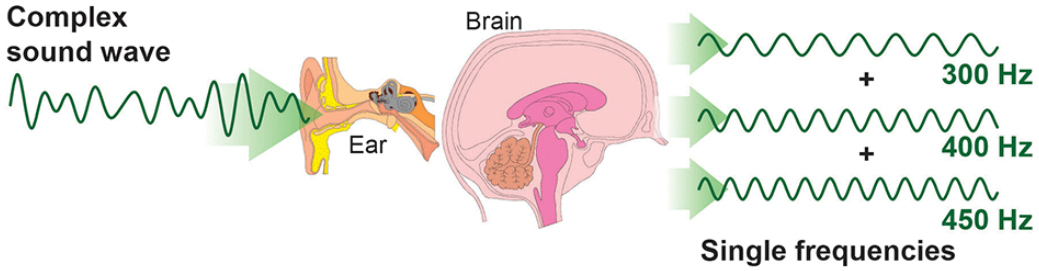


Figure 3-11: Sound human brain understanding [**?**]

3.3 Motor Gestures

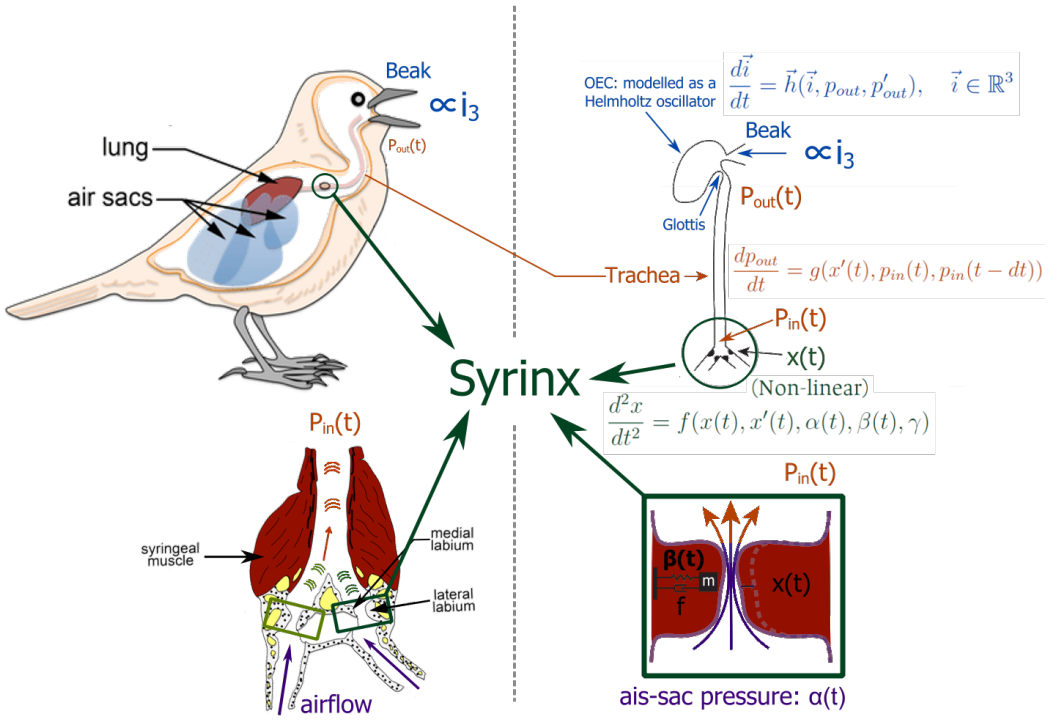


Figure 3-12: Schematized view of a dynamical systems model describing syringeal labial dynamics and tracheal vocal-tract filtering. The syringeal membrane was modeled as a mass (m) with external pressure (α) and a restitution (spring) constant (β). Here: γ, time constant; r , reflection coefficient of the trachea; $T = L/C$, propagation time along trachea; v, proportional to the mean velocity of the flow; $y = x'$, velocity. Images taken and adapted from [**?**] and [**?**].

The most complete and tested physical model to simulate the birds sound production is the **motor gestures model**. It has been developing in the Dynamical System Lab, lead by professor G. Mindlin of the National University of Buenos Aires (UBA). The model simulate the bird sound production system by modeling the syrinx, trachea, Oro-Oesophageal Cavity (OEC), glottis, and beak with ordinary differential equations (ODEs).

3.3.1 Timeline

The described bird's vocalization model was developed based on the the human vocals folds literature models. The first model of human vocal folds date at 1988 and was created by Titze, [**?**], where he studied how the muscles movement generate a pressure wave that yields to a vocalization due to the muscles oscillations.

This model consider the vocal folds as a simple mass-spring oscillator, a second order differential equation, with a nonlinear driven force

$$m\xi'' + b\xi' + k\xi = f(\xi(t), \xi'(t), t)$$

where ξ, ξ', and ξ'' are the wall vocal fold position, velocity and acceleration, respectively, m the mass, k the stiffness, b the damping, f the driving force, and t the time. The study of the dynamical behavior of the Titze model shows that the oscillation occurrence depends of the driving force f and the velocity term, which involves the damping constant and vocal folds walls velocity. Since this model is basic it has great interpretability and has been used as a toy model to simulate human and animal vocalizations.

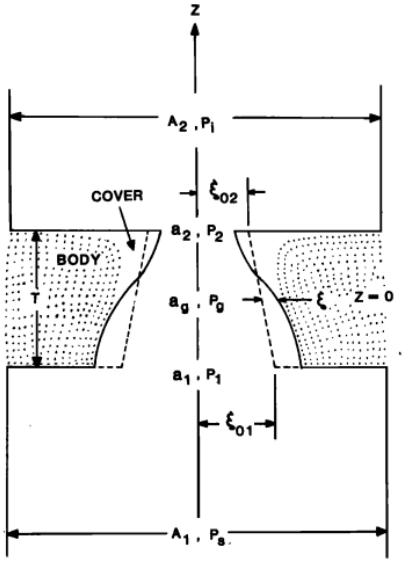


Figure 3-13: Frontal section of body-cover model, vocal folds, used for small-oscillation analysis [**?**], a toy model.

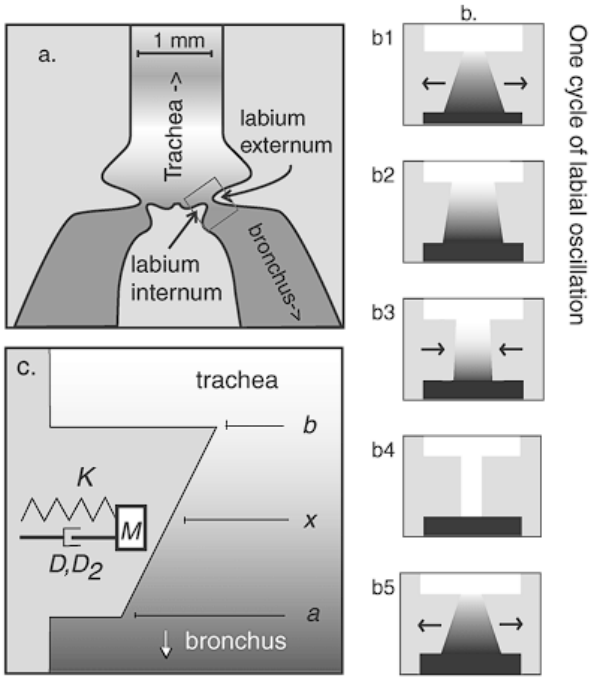


Figure 3-14: Syrinx behavior illustration and labial dynamics. The panel (a) illustrates the present organ actors and their shape, (b) shows a representation of labial position in an oscillation cycle, and (c) is a diagram with the model terms. [**?**]

Using the Titze model and studying its dynamical behavior, professor Mindlin proposed the first model to simulate birds sound production, **motor gestures for birdsongs**, making possible to create simple synthetic birds vocalizations (birdsongs). This model of bird's sound production was published at 2001 [**?**] taking advantage of the bifurcation theory ideas and applied them to the Titze model. Although Titze model was developed for human vocal folds, experimental evidence [**?**] confirm the similarity between the vocal folds and birds labial oscillations. This model successfully generate approximate synthetic birdsongs by studying some control parameters for a Titze modified model, two driving parameters.

Since the article with the proposed model just make a qualitative comparison, the next step will be test the model by comparing its result (synthetic syllables) with real birdsongs. At the following year, 2001, a model improvement was published. There, the motor gesture model was improved by using more advanced mathematics analysis, generating a simpler EDO system for the labia position, and by making a quantitative comparison, reproducing experimental recorded birdsongs of the Chingolo sparrow (Zonotrichia capensis or copeton) [**?**]. This article shown how bird syllables can be generated by simple curves (paths) of the model parameters (three dimensional parameter space).

In the same year, another model improvement was made: instead of use recorded birdsongs the neural subpopulation activity in the robust nucleus of the archi-striatum (RA) was used. Then, the new proposed model allow to us to study how some neuron brain activity generates labial oscillations, certain connectivity architectures in the RA give rise to a wide range of different vocalizations under simple excitatory instructions. [**?**]

For the following years, the model was studied with more advanced mathematics, using dynam-

ical systems theory to study where and how the labial oscillations appear, and even a validation experiment was carried out. At 2003 an article present new experimental support data to validate the motor gestures model. Here, the model uses as control parameters some functions whose time dependence comes from recordings of muscle activities and air sac pressure. In addition, the birdsong was simultaneously recorded to compare it with the synthetic birdsong. This time the model generates recognizable birdsongs and tested some predictions concerning to the relative levels of activity in the syrinx muscles. [**?**] Not enough of experiments, two years later using the motor gestures model and parameters an electronic syrinx was developed and tested [**?**]

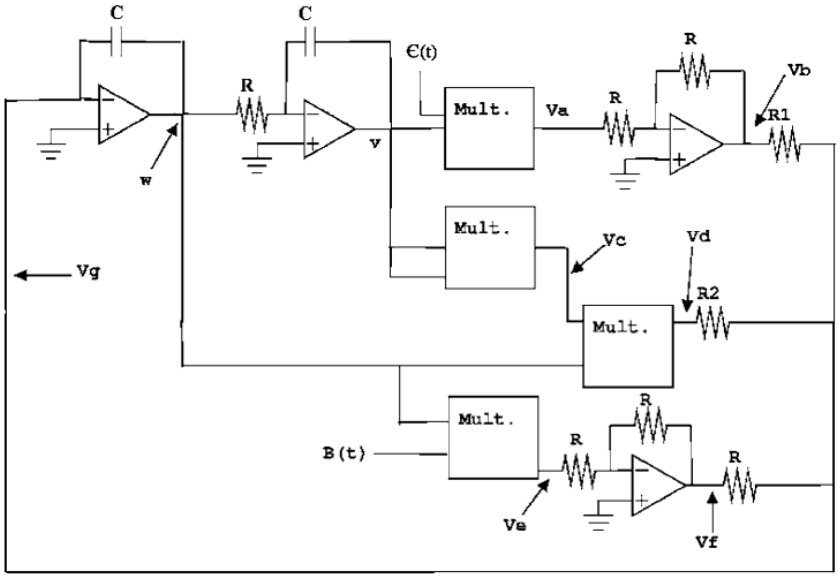


Figure 3-15: Circuit of the electronic syrinx. This circuit takes into account the the syringial labial position by a dynamical model. Two control parameters are required to drive the system into the oscillation region in the parameters space [**?**]

At this point, the model exhibits great accuracy to generate synthetic birdsongs by just modeling the syrinx as the unique actor in the bird sound production. Keeping the model improvement a new sound production organ actor is added to the model in the same year: the vocal tract (trachea), which is the responsible of the sound filtering [**?**]. Moreover, since the model has new equations and variables it is necessary to study its dynamical behavior to understand what is the dynamical origin of the spectrally birdsongs richness. This articles shown that the spectral content carries a strong signature of the intrinsic dynamics of the sound source, here the sound spectral index (SCI) is introduced to compare the syllables spectral content.[**?**]

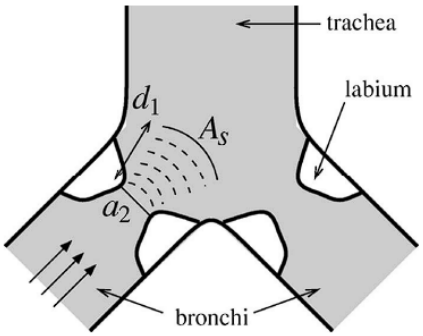


Figure 3-16: Coupling between syrinx labia and vocal tract (trachea). Schematic central cross section. The air sac pressure comes from the bronchis and is modulated by the airflow-induced oscillation of the labia, that then injects a sound pressure wave into the base of the trachea. [**?**]

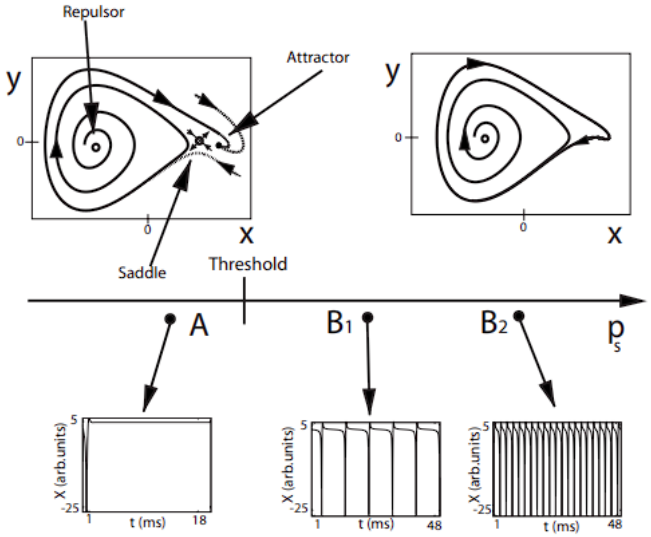


Figure 3-17: Bifurcation diagram for the two possible model system dynamics. When the air sac pressure p_s is lower than a threshold the diagram has three fixed points but not labial oscillation occurs, while when the air sac pressure crosses the bifurcation line oscillations born, for values around the threshold the oscillation has zero frequency but for far way values the oscillation starts to be tonal. [**?**]

Complementary to all the research made, in the same year a birdsongs book is published: **The physics of birdsongs** [**?**], that summarize and compile all the information about bird and their sound production models. The book is self-contained, it includes the necessary physical concepts and mathematical background, even it has a chapter about the brain rule in the sound production process. Along with this book, four years later, at 2009, a supporting review of the birdsong models is published in Scholarpedia [**?**].

The final actors to be added are the OEC cavity, glottis, and beak. This organs were include until 2011, using a new bifurcation (Bogdanov-Takens), the glottis is modeled as the neck of the Helmholtz resonator representing the OEC that interacts with the atmosphere through the beak. In this articles the physiological instructions from Zebra Finch[2] song are reconstructed, the birdsongs are qualitative compare by computing the SCI and pitch differences, generating highly realistic synthetic songs.

Finally, the complete model is published and well describe in the articles [**?**] and [**?**]. In both articles the dynamical behavior is studied and the Zebra Finch control parameters are computed,

[2]This is song is used since it is widely studied by ornithologist

they find the space parameters paths that better reproduces the zebra finch songs.

Next to the complete model publications, the model has been explored with new ideas to be upgraded as the vortex inclusion in the system dynamics [**?**] or use machine learning tools to analyze and create surrogate birdsongs [**?**, **?**, **?**].

3.3.2 Current Model

The previous chapter demonstrated how good is the model, along many years it have been tested and upgraded by the DSL laboratory developing a complete functional vocal bird organ, moreover the syrinx, through a physical model named motor gestures. It consist of simulate the syrinx, trachea, glottis, OEC and beak [**?**].

- **Syrinx**

  The main character of this model is the syrinx. It is modeled as a mass below the act of an elastic force and a nonlinear damping, defined by the bifurcation theory such that the syrinx movement generate oscillations. Hence, the labia position will satisfy the following second order non-linear differential equation

$$\frac{dx^2}{dt^2} = \gamma^2[-\alpha(t) - \beta(t)x + x^2 - x^3] - \gamma(1+x)x\frac{dx}{dy}$$

  making a simple substitution, this equation can be write as set of two first order linear differential equations

$$\begin{aligned} \frac{dx}{dt} &= y \\ \frac{dy}{dt} &= \gamma^2[-\alpha(t) - \beta(t)x + x^2 - x^3] - \gamma(1+x)xy \end{aligned} \tag{3-1}$$

  The syrinx is modeled with the Bogdanov Takens bifurcation, subsection 2.4.5, but this is not the only possibility. Another approach is to use a different bifurcation, use another damping function, as was tried in the first model publication [**?**].

  To study the ODEs bifurcations let us start to calculating the equilibrium points. This points occurs when the variables derivatives are zero

$$\begin{aligned} \frac{dx}{dt} &= y = 0 \\ \frac{dy}{dt} &= \gamma^2[-\alpha(t) - \beta(t)x + x^2 - x^3] - \gamma(1+x)xy = 0 \\ &\Rightarrow \quad \gamma^2[-\alpha(t) - \beta(t)x + x^2 - x^3] = 0 \end{aligned}$$

Then, the derivatives vanishes when the labial velocity is zero (curve $y = 0$) and the air sac pressure and labial tension satisfies the previous relation that can be solved for α

$$\alpha(t) = -\beta(t)x + x^2 - x^3 \tag{3-2}$$

Thinking geometrically, this is equivalent to find if and where the two curves intersect: the horizontal line define by the air sac pressure $g(x) = \alpha$ and the labial position condition define by the cubic polynomial $f(x) = -\beta x + x^2 - x^3$. In order to get oscillations the tangent of the functions must be equal and the control parameters must satisfies the relation $0 = -\beta + 2x - 3x^2$ which give a solution for the parameter β that later are used for calculate the bifurcations curves present in the Takens Bogdanov, two saddle-node and one hotf bifurcations curves.

- **Trachea (windpipe)**

The trachea acts as a filter and is modeled as a tube with one end opened [**?**], the end connected to the syrinx, and the other one closed, the end that is connected to the glottis. The output pressure of the trachea satisfies a delayed differential equation of the input pressure, they are proportional directly. Furthermore, since the wave can be absorbed and reflected by the trachea a reflection coefficient is defined r such that output pressure is

$$p_i(t) = Ay(t) + p_{back}\left(t - \frac{L}{c}\right) \quad ; \quad p_{back}(t) = -rp_i\left(t - \frac{L}{c}\right) \tag{3-3}$$

$$p_{out}(t) = (1 - r)p_i\left(t - \frac{L}{c}\right) \tag{3-4}$$

$$\frac{dp_{out}}{dt}(t) = \frac{p_{out}(t) - p_{out}(t - dt)}{dt} \tag{3-5}$$

here p_{back} is the backward pressure, L the trachea length, and c the velocity of the sound in air. The last equation is the the approximation by backward difference of the output pressure derivative.

- **Beak, Glottis, and Oro-Oesophageal Cavity (OEC)**

Previous works showed how a simple circuit can simulate the filtering audio process done by the birds [**?**, **?**] : the glottis is modeled with a resistor; the OEC as a Helmholtz resonator with an inductance and a resistor; and the beak as an inductance in series with a resistor. Since the the filtering circuit, Figure **3-20**, is analog to the filtering bird process, the current is equivalent to the pressure and hence there will be an analog Kirchhoff voltage law for the pressure

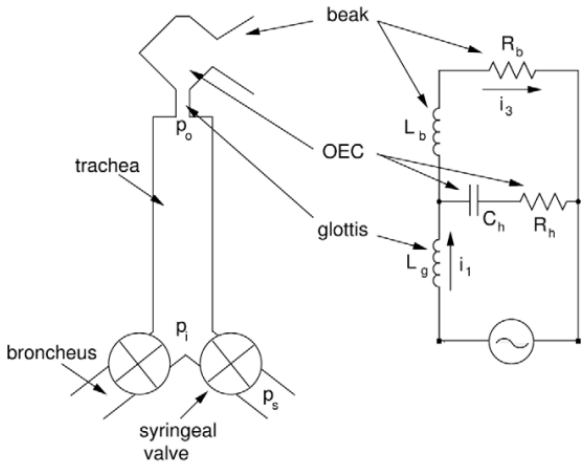


Figure 3-18: [?]

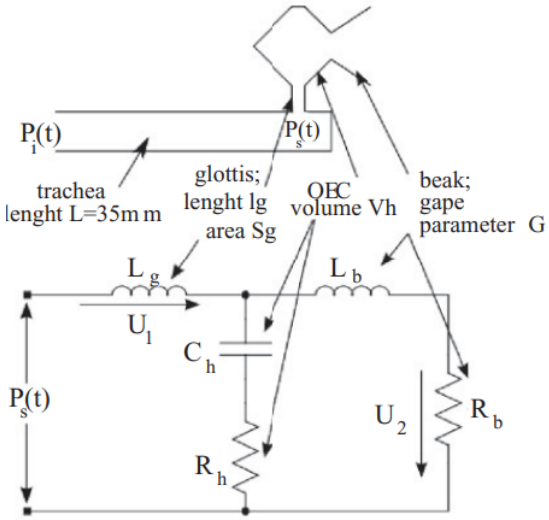


Figure 3-19: [?]

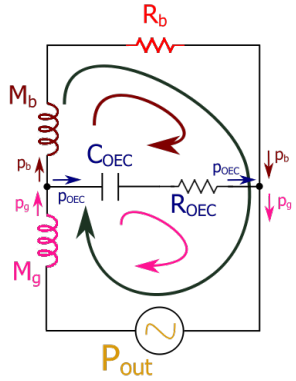


Figure 3-20: Analog voltage Kirchhoff loops.

using an analog acoustical Kirchhoff voltage law in all the loops

$$p_{out} = M_g p'_g + \frac{1}{C_{OEC}} \int p_{OEC} dt + R_{OEC}\, p_{OEC} \tag{3-6}$$

$$M_b p'_b + R_b p_b + \frac{1}{C_{OEC}} \int p_{OEC} dt + R_{OEC}\, p_{OEC} = 0 \tag{3-7}$$

$$p_{out} = M_g p'_g + M_b p'_b + R_b p_b \tag{3-8}$$

$$p_g = p_b + p_{OEC} \tag{3-9}$$

taking the derivative respect to time

$$\begin{aligned} M_g p''_g + \frac{1}{C_{OEC}}\, p_{OEC} + R_{OEC}\, p'_{OEC} - p'_{out} = 0 \\ M_b p''_b + R_b p'_b + \frac{1}{C_{OEC}}\, p_{OEC} + R_{OEC}\, p'_{OEC} = 0 \\ p'_{out} = M_g p''_g + M_b p''_b + R_b p'_b \end{aligned} \tag{3-10}$$

making a change of variable and solving the last equation for p'_b

$$p'_g = dp_g, \qquad p''_g = dp'_g \tag{3-11}$$

$$p'_b = \frac{1}{M_b}\left[p_{out} - R_b p_b - M_g p'_g\right] = \frac{1}{M_b}\left[p_{out} - R_b p_b - M_g dp_g\right] \tag{3-12}$$

changing the pink equation to

$$M_g p''_g + \frac{1}{C_{OEC}}\left(p_g - p_b\right) + R_{OEC}\left(p'_g - p'_b\right) - p'_{out} = 0$$

$$M_g dp'_g = -\frac{1}{C_{OEC}}\, p_g + \frac{1}{C_{OEC}}\, p_b - R_{OEC}\, dp_g + R_{OEC}\frac{1}{M_b}\left[p_{out} - R_b p_b - M_g dp_g\right] + p'_{out}$$

$$M_g dp'_g = -\frac{1}{C_{OEC}}\, p_g + \left(\frac{1}{C_{OEC}} - \frac{R_{OEC}R_b}{M_b}\right) p_b - R_{OEC}\left(1 + \frac{M_g}{M_b}\right) dp_g + \frac{R_{OEC}}{M_b} p_{out} + p'_{out}$$

solving for dp'_g

$$dp'_g = -\frac{1}{C_{OEC}M_g}\, p_g + \frac{1}{M_g}\left(\frac{1}{C_{OEC}} - \frac{R_{OEC}R_b}{M_b}\right) p_b - R_{OEC}\left(\frac{1}{M_g} + \frac{1}{M_b}\right) dp_g + \frac{R_{OEC}}{M_b M_g} p_{out} + \frac{1}{M_g} p'_{out} \tag{3-13}$$

Hence, the set of first order differential equations that simulate the whole filtering bird organ are defined by the equations(3-11), (3-12) and (3-13)

$$\begin{aligned}
\frac{d}{dt}p_g &= p'_g = dp_g \\
\frac{d}{dt}dp_g = dp'_g &= -\frac{1}{C_{OEC}M_g}p_g - R_{OEC}\left(\frac{1}{M_b} + \frac{1}{M_g}\right) dp_g + \frac{1}{M_g}\left(\frac{1}{C_{OEC}} + \frac{R_{OEC}R_b}{M_b}\right) p_b \\
&\quad + \frac{1}{M_g}\frac{dp_{out}}{dt} + \frac{R_{OEC}}{M_g M_b} p_{out} \\
\frac{d}{dt}p_b = p'_b &= -\frac{M_g}{M_b}p'_g - \frac{R_b}{M_b}p_b + \frac{1}{M_b}p_{out}
\end{aligned} \tag{3-14}$$

with M, R, and C denoting the Intertance, Resistance, and Compilance, acoustical analogs of Impedance L, Resistance, and Capacitance, electrical variables, respectively. The length and area of an element a are l_a and S_a, respectively, which in our model stand for the beak (b), the glottis (g), and the OEC volume (OEC) [**?**, **?**].

Impedance	Inertance
L	$M = \frac{\rho_0 l_a}{S_a}$
Resistance	Resistance
R	$R = \frac{\rho_0 c k^2}{2\pi}$
Capacitance	Compliance
C	$C = \frac{V_b}{\rho_0 c^2}$

Figure 3-21: Acoustic and electrical analogs [**?**]

reference [**?**], [**?**], [**?**],

3.3.3 State of Art

Currently, the model is being used to train machine learning (ML) algorithms to identify and generate bird songs, and even its dynamical behavior is still being studied with ML ideas.

The model automatizing has been discussed from many years ago. The first article that proposed a solution method to find the parameters automatically was published in 2015, [**?**], where the idea is to produces a grid for the control parameters and calculate at each node of the grid the Spectral Content Index (SCI) and the Fundamental Frequency (FF or pitch) to find what are the control parameters values, the motor gesture (synthetic audio), that generates a good approximation of the real birdsong, this grid computation is done previously and is stored in data files.

New article research shows how neural networks can help to solve the problem of automatizing parameters using as input data the neural population activity recorded from electrode arrays implanted in the premotor nucleus HVC [**?**].

The present work presents, describes and evaluate a computational physical model to automatize the motor gestures generation from audio data recorded. In addition, the model is implemented as python package, using POO thinking to the implementation and Github to storage, such that make it easily to use, to share and develop.

3.4 Automatizing

The final step for the model is make it automatic an portable: that any person can download an efficient implementation of the model as package, easy to use and fast to execute. This objective can be achieve by designing, implementing an evaluating a computational physical model that given an audio signal it

- Characterize the audio signal by computing its fundamental frequency (FF), also called pitch, and its tempo-spectral features.
- Generate a synthetic syllable from the input signal and some model parameters (α, β, γ).
- Compare the real and synthetic signals.
- Find the best optimal parameters values that better simulate the real syllable (birdsong).
- Split an audio song into single syllables.
- Write an audio file with the synthetic syllable (song).

The final result is an available online package that generates synthetic birdsongs from real audio data, by using recorded birdsongs in audio wav format. This package allows not just to generate the most similar synthetic birdsong, it also offers the possibility to generate a huge amount of syllables related to the real audio. Moreover, the model implementation allows to change the bifurcation, using symbolic calculus, and explore the physical model behavior, by varying the system variables values (α, β, γ).

3.4.1 Optimization Problem

The goal of this work is to find automatically some optimal parameters from experimental samples, real audios. This can be formulated as an inverse problem with a white box[3], the goal is to find the optimal parameters that causes the measured observations using the white box model. Since the problem is not linear the optimal parameters may not be unique.

This kind of problems have been studied from many years ago and there are many theories created to solve them, one of these theories is the numerical optimization that offers a new tool to solve a wide variety of inverse problems by defining them as a minimization (or maximization) optimization problem.

Using the numerical optimization theory, the motors gestures automatizing problem can be formulated as the following minimization that dependes of three control parameters: air-sac pressure α, labial tension β, and a constant time γ, two vector arrays and one real number respectively.

$$\begin{aligned} \min_{\gamma\in\mathbb{R},\ \alpha,\beta\in\mathbb{R}^n} \quad & ||\hat{SCI}_{real} - \hat{SCI}_{synt}(\gamma,\alpha,\beta)||_2 + ||(\hat{FF}_{real} - \hat{FF}_{synt}(\gamma,\alpha,\beta)||_2 \\ & - corr(FC_{real}, FC_{synt}(\gamma,\alpha,\beta)) \\ \text{subject to} \quad & \gamma\in\Omega_\gamma, \quad \beta\in\Omega_\beta, \quad \alpha\in\Omega_\alpha \end{aligned} \tag{3-15}$$

Here $corr$ is the acoustic dissimilarity between two spectral coefficients sets, FC the Fourier spectral coefficients, $\hat{FF}$ the dimensionless fundamental frequency (divide by 1 kHz and its size), $\hat{SCI}$ the spectral content index divide by its size, and Ω_i represents the feasible region for the variable i.

This minimization problem is nonlinear, then nonlinear optimization algorithms are required, and unconstrained. However, the Jacobian and Hessian matrix calculations are not simple and require a deeper study in order to be able to implement better algorithms and be ensure of the convergence to a solution. Further techniques can also involve constraints, in this problem the most appropriated constraints are the bifurcations curves of the labia position.

[3] physical model where the behavior is well known, motor gestures

4 Methods and Methodology

The model is implemented in the general-purpose programming language Python 3 and hosted in GitHub. You can find the repository at birdsongs.

Today, Python is one of the most widely used programming languages in the world, both by industry and academia. Not only because is easy to learn and write, or because it is popular and used in many fields: websites and software developing, task automation, data analysis, data visualization, server management, or math and scientific computing; it is a highly reliable and efficient language, which allows to user to share and develop code in an easy way due to the supporting of multiple programming paradigms including structured, object-oriented, and functional programming.

4.1 Programming Object Oriented (POO)

POO is more than a way to think about code, it relates to the way code is created, the software architecture, such that enables flexibility through modular design and maximum reusability. [**?**]

This programming paradigm offers the user the advantage to easily understanding the operation of a program by gathering data (attribute) and its behavior (method) in a single bundle called an **object**. Some of the advantages are:

- Reusability
- Testing and debugging
- Flexibility
- Speed Up

The major disadvantage is that it is a more arduous and labor-intensive process. [**?**].

4.1.1 Python Objects

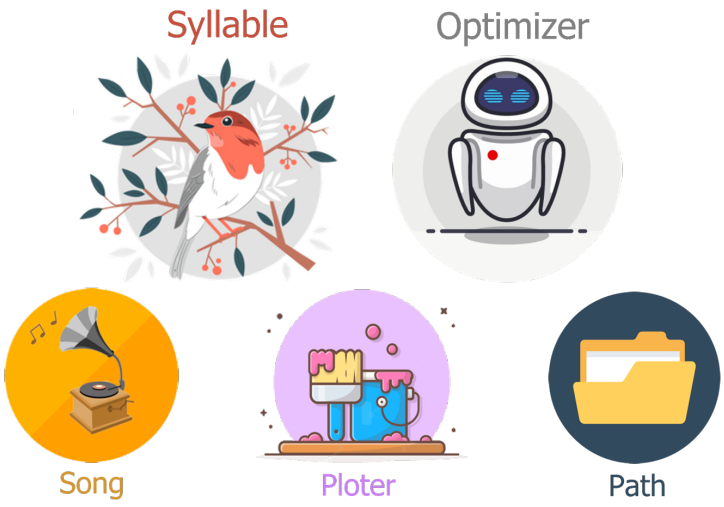


Figure 4-1: Objects created to model, generate, solve, and visualize synthetic and real birdsongs.

4.2 Model Implementation

4.2.1 Audio Processing

The signal processing techniques used here are

- Fundamental frequency computing.
- Fourier transform.
- Signal falterings.

they were implemented using the signal processing theory and librosa package [**?**, **?**].

4.2.2 Methodology

In order to make the code shareable and public the implementation is made,

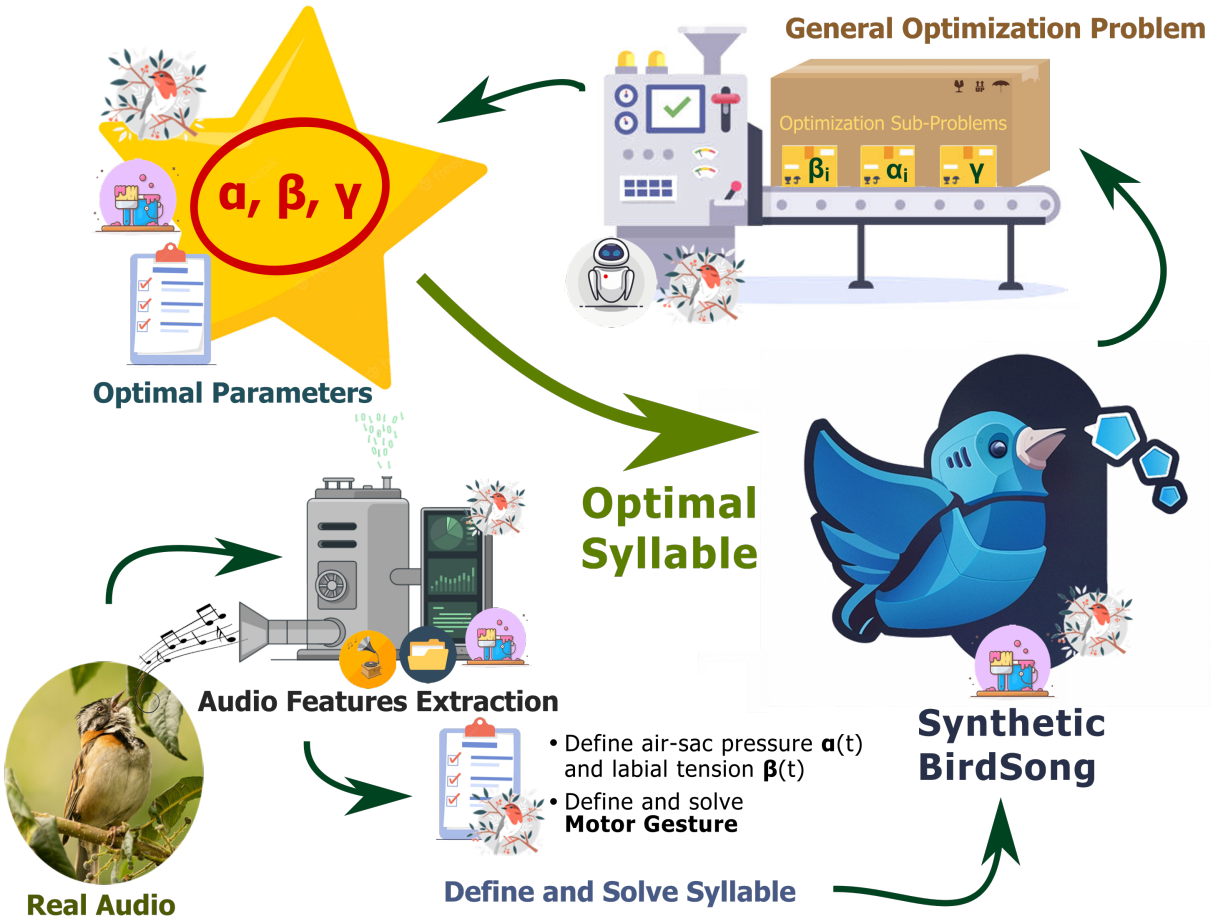


Figure 4-2: Schematic programming methodology used to implement and solve the problem.

4.2.3 Numerical Optimization

General Problem

As was discussed in the previous chapter, the automatization of the model can be summarized to solve the optimization problem defined by the equation (3-15). The major issue of this problem is the parameters space, the optimal solution is in a high dimensional space $(2n + 1)$ and finding it is a difficult task, and exhaustive and time consuming process since n is usually a huge number.

One solution to this drawback is to solve auxiliary optimization problems, each control parameter has its own optimization problem, with lower dimensions. Although this is a good approach to solving the problem, it is still a huge process because the dimensions of the parameters n is still huge[1].

[1]the size of the control parameters is the same size of the audio signal

The solution is to parameterize the control parameters in terms of three or two coefficients for each vector parameter, whichreduces the dimension problem from $2n+1$ to at most $2(3)+1$. With these ideas the general optimization problem can be formulated as

$$\begin{aligned} \min_{\gamma\in\mathbb{R},\, a,b\in\mathbb{R}^3} \quad & ||\hat{SCI}_{real} - \hat{SCI}_{synt}(\gamma,a,b)||_2 + ||(\hat{FF}_{real} - \hat{FF}_{synt}(\gamma,a,b)||_2 \\ \text{subject to} \quad & \gamma\in\Omega_\gamma, \quad b\in\Omega_b, \quad a\in\Omega_a \end{aligned} \tag{4-1}$$

with $\Omega_\gamma = [1000, 100000]$, $\Omega_a = [0.01, 0.25]\times[-2,2]\times[0,2]$ and $\Omega_b = [-1, 0.5]\times[0.2,2]\times[0,3]$ the feasible regions for each variable. In order to get an objective function dimensionless the following two variables are define

$$\hat{SCI} := \frac{SCI}{dim(SCI)}, \qquad \hat{FF} := \frac{1}{dim(FF)}\frac{FF}{1\ KHz}$$

where $dim()$ is the dimension the corresponding vector.

To reduce the dimensionality problem, the air-sac pressure and labial tension over the time are parameterized in term of at most three coefficients, that depends of time, as follows

$$\alpha(t) = a_0 + a_1 t + a_2 t^2, \tag{4-2}$$

$$\beta(t) = \begin{cases} b_0 + b_1\left(\frac{FF_{real}}{10^4}\right) + b_2\left(\frac{FF_{real}}{10^4}\right)^2, & \text{id=syllable} \\ b_0 + b_1 t + b_2 t^2, & \text{id=chunck} \end{cases} \tag{4-3}$$

with the same time duration (T) as the input signal, $t\in[0,T]$.

Optimization Sub-problems

The general problem is computationally expensive since it still depends on several variables. Although solving the problem all at once is the ideal method, a better approach is to divide it into three auxiliary problems, one optimization problem for each control parameter.

- **Optimal Time Scaling Constant** (γ)

  The time scaling coefficient parameter (γ) should be the same for all syllables of the same song, in general a bird has the same γ value for all its syllables. This parameters modifies the spectral content, increasing or decreasing it, so to find its optimal value the following optimization problem must be solved

$$\begin{aligned} \min_{\gamma\in\mathbb{R}} \quad & ||\hat{SCI}_{real} - \hat{SCI}_{synt}(\gamma)||_2 + ||\hat{FF}_{real} - \hat{FF}_{synt}(\gamma)||_2 \\ \text{subject to} \quad & \gamma\in\Omega_\gamma = [10000, 100000] \end{aligned} \tag{4-4}$$

  In this auxiliary problem the other control parameters (α and β) are taken constant but ensuring they are in region of oscillations defined by the bifurcation theory, the feasible region is defined by the analyze of parameters behavior of previous works.

- **Optimal Air-Sac Pressure (α) Coefficients**

  The parametric coefficients of the air-sac pressure (α) are calculated by solving the following maximization problem

$$\begin{aligned} \max_{a \in \mathbb{R}^3} \quad & corr(real, synthetic(a)) \\ \text{subject to} \quad & a \in \Omega_a \end{aligned} \tag{4-5}$$

  This is done by computing the acoustic dissimilarity between the coefficients of the real and synthetic spectrums, this set of coefficients has the information of the harmonic of the audio signal. Then, the dissimilarity correlation is an appropriated objective function to solve the search for the optimal parameters for the air-sac pressure, remember that this control parameter impacts the intensities and distances of the harmonics of the signal.

  As every maximization problem is equivalent to solve the a minimization problem, instead of solving the problem (4-5) we will solve the following minimization problem

$$\begin{aligned} \min_{a \in \mathbb{R}^3} \quad & -corr(real, synthetic(a)) \\ \text{subject to } & a \in \Omega_a \end{aligned} \tag{4-6}$$

  where the feasible region, $\Omega_a = [0, 0.25] \times [-2, 2] \times [0, 2]$, is defining by the study and exploration of the model and its bifurcation theory.

- **Optimal Labia Tension (β) Coefficients**

  The last step is to find the parametric coefficients of the labial tension b_i. Since the labial tension is the key player in the sound production of birds, as it modulates the air at certain frequencies, the fundamental frequency is highly dependent on this control parameter. Therefore, a good objective function should include this spectral variable.

  The optimization problem defined to find the optimal parameters of the labial tension is

$$\begin{aligned} \min_{b \in \mathbb{R}^3} \quad & ||FF_{real} - FF_{synt}(b)|| \\ \text{subject to} \quad & b \in \Omega_b \end{aligned} \tag{4-7}$$

  where the feasible region is $\Omega_b = [-1, 0.5] \times [0.2, 2] \times [0, 3]$

At present, there are many algorithms and packages created to solve any type of optimization problem, from linear to nonlinear[2]. Because the model was implemented in Python and after a literature review, the most appropriated library to solve the optimization problems presented in this work is the Non-Linear Least-Squares Minimization and Curve-Fitting for Python (lmfit [**?**]), which is both a powerful and easy to use.

[2] Soome lirabries as CasaADI, Matlab optimization toolbox, of lmfit [**?**]

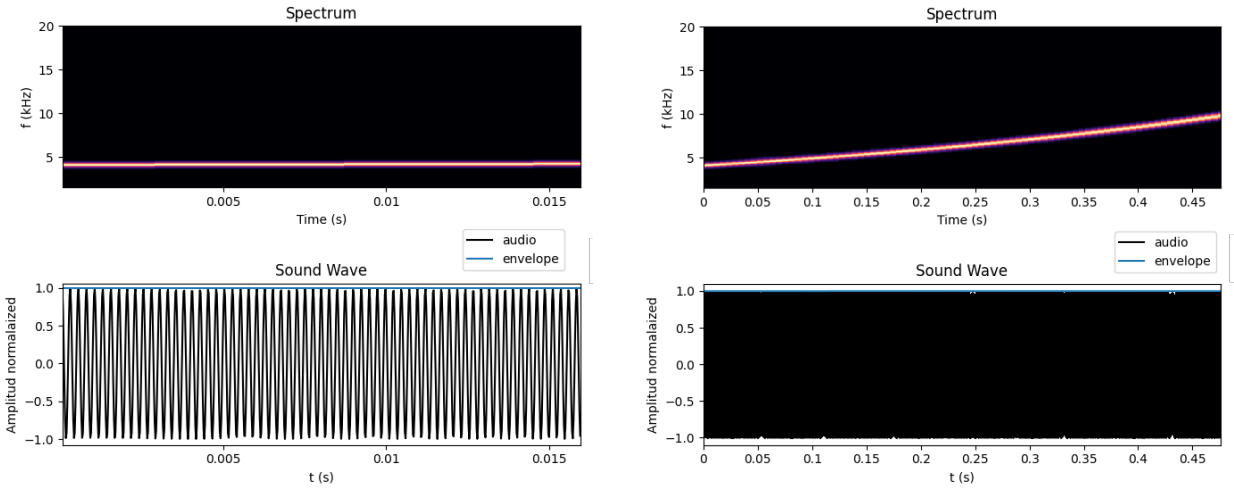
5 Results and Discussion

5.1 Model Implementation

In this chapter the result are presented and discussed. Since the main goal of the work is the packing of the model, the extensive and statistical evaluation of the model is missing.

5.1.1 Motor Gestures

The python implementation of the model allows to easily and quickly explore the sound production model of birds. Using the created objects: syllable, optimizer, birdsongs, paths, and ploter, **??**, it is possible to explore and explain how the value parameters impact the output sound (synthetic syllable/birdsong). This is important for a parameters sensitive analysis.



(a) Little part of the input signal, chirp.

(b) Whole input signal

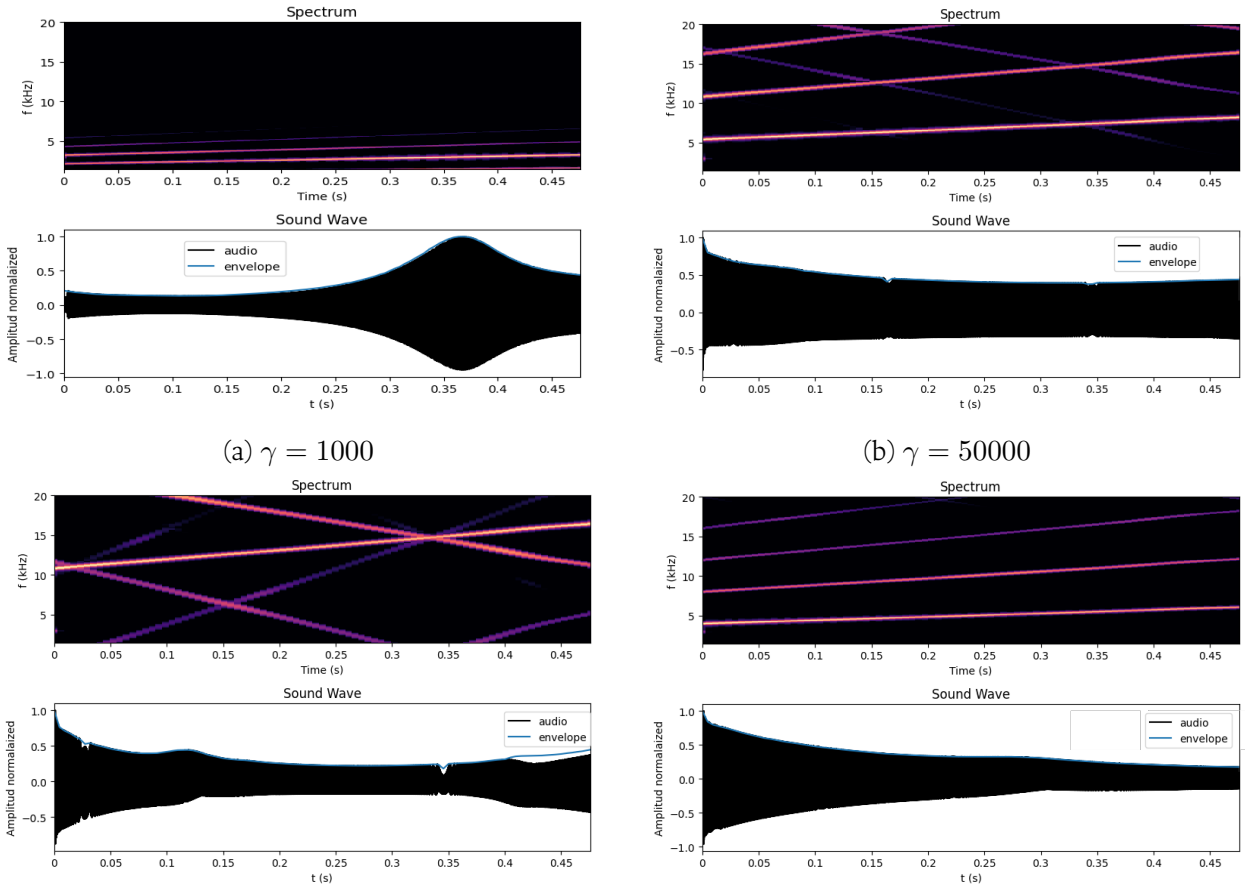
Figure 5-1: Input chirp signal wave form and spectrogram.

As an example consider a pure synthetic case, the external trachea pressure (envelope of the signal function $A = e$, equation (3-3)) is a simple signal (such as a chirp or tone), implying nether audio file or initial coefficient parameters are required. Since this is a simple function, it can be approximated as a quadratic or linear function, then the parameters coefficients can be better understood: the constant coefficient represents a vertical shift for the parameters curves, the

linear coefficient gives both vertical and horizontal displacements, and the quadratic coefficient modifies the curvature of the function. In fact, the parameterization of the parameters not only minimizes dimension of the search space, but also provides interpretability for each coefficient.

Note that it is possible to define any input signal for the external pressure and explore its impact on the model. Interesting cases are pulses, clicks, tones, or chirp signals, which are easy to experimentally reproduce.

First we will study the meaning and impact of the scalar time constant γ for the synthetic syllable. For this purpose let us consider some coefficient parameters and three values for $\gamma = 1000, 50000$, and 100000, Figure 5-2 and the following parameters coefficients $a = [0.11, 0.05, 0]$ and $b = [-0.1, 1, 0]$.



(a) $\gamma = 1000$

(b) $\gamma = 50000$

(c) $\gamma = 100000$

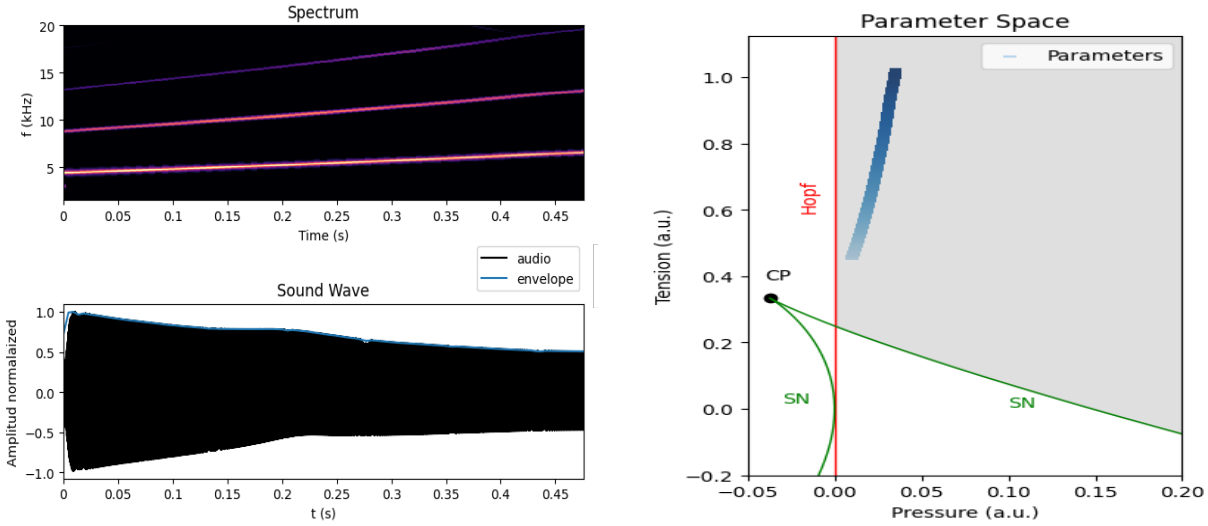
(d) $\gamma = 37000$

Figure 5-2: Spectrogram of output signal varying the time constant scale.

When γ takes low values the frequency range is small and generates a lower fundamental frequency which may have harmonics (it depends on how far away the sac pressure and labia ten-

sion parameters are from the bifurcations, the Hopf bifurcation does not generate harmonics while the Saddle-Nodes do). In contrast, when γ takes high values the frequency range increase and many harmonics are present. The optimal values for zonotrichia are those in the middle of the entire range, values between 30000-50000.

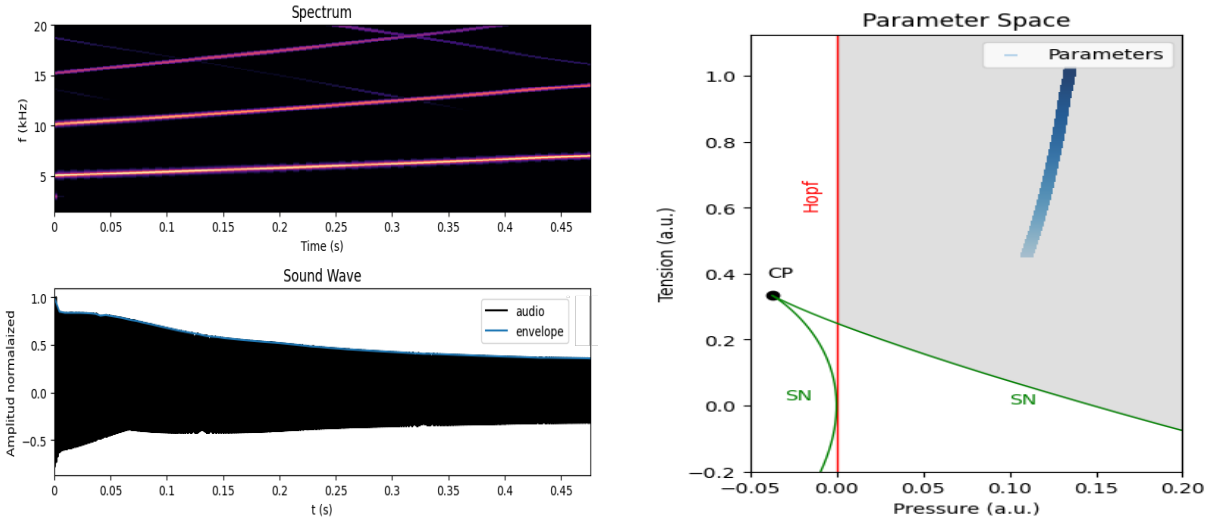
Now let us explore the impact of the air pressure on the output signal, more so on the fundamental frequency of the synthetic birdsong which is equivalent to exploring different intensities of air pressure from the the bird's lungs entering the syrinx. Let us consider the same coefficient parameters set defined above.



(a) Wave and spectrogram

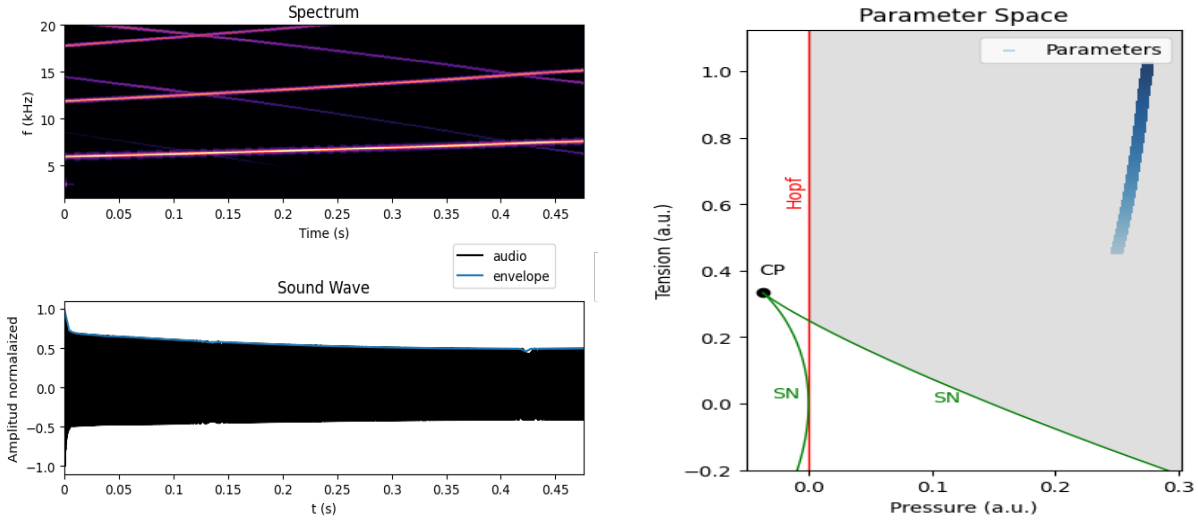
(b) Parameters curve an bifurcations

Figure 5-3: Low value $a_0 = 0.01$.



(a) Wave and spectrogram (b) Parameters curve an bifurcations

Figure 5-4: Middle value $a_0 = 0.11$.



(a) Wave and spectrogram (b) Parameters curve an bifurcations

Figure 5-5: High value $a_0 = 1.25$.

Similarly the parameters space of the labial pressure can be explore. If the audio is less than 10 milliseconds it is defined as a chunck and the labia initial curve is at most a quadratic function defined by its parameters, while if the audio is greater than 10 milliseconds it is defined as a syllable having as initial curve of the labia tension the same fundamental frequency curve but rescaled and dimensionless.

5.1.2 Syllable

The main main bird of study is the Zonotrichia Capensis. They are typical birds in Latin America and specially in Colombia, they are present almost in all the country where we have recorded and published many birdsongs. To study a birdsong, a composition of several syllables or chunck, of any bird it is necessary split the song by syllables and study them one by. Let us study the following Zonotricha birdsong as a syllables and the thrilled part as a chunck, generally this bird has the same form of pitches curves but in the major of cases this pitches has simple shapes.

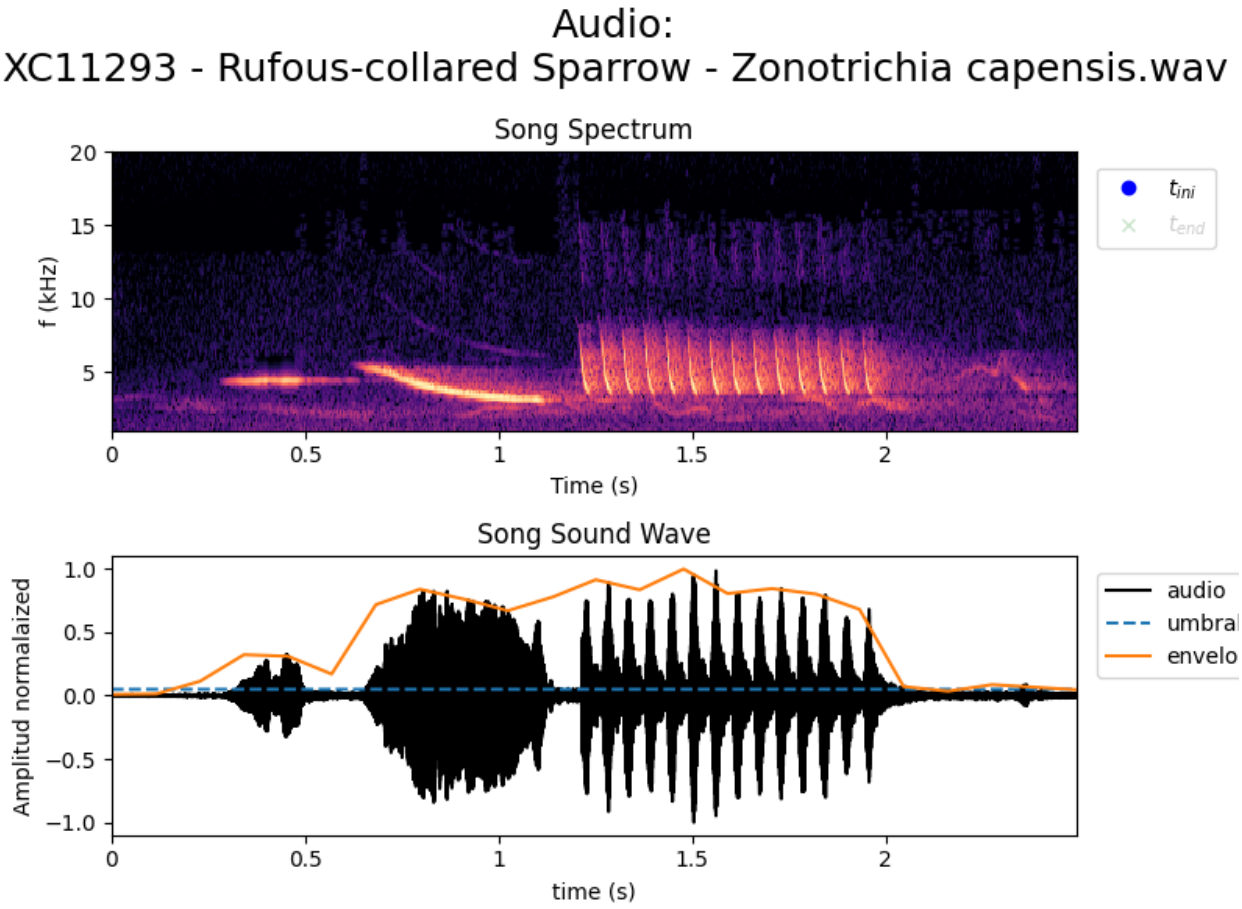


Figure 5-6: Zonotrichia capensis birdsong sample. It is an audio with noise and three sections of birdsongs: two clear and simple syllables and one series of thrilled syllables. The envelope can be improved with higher Nt object values

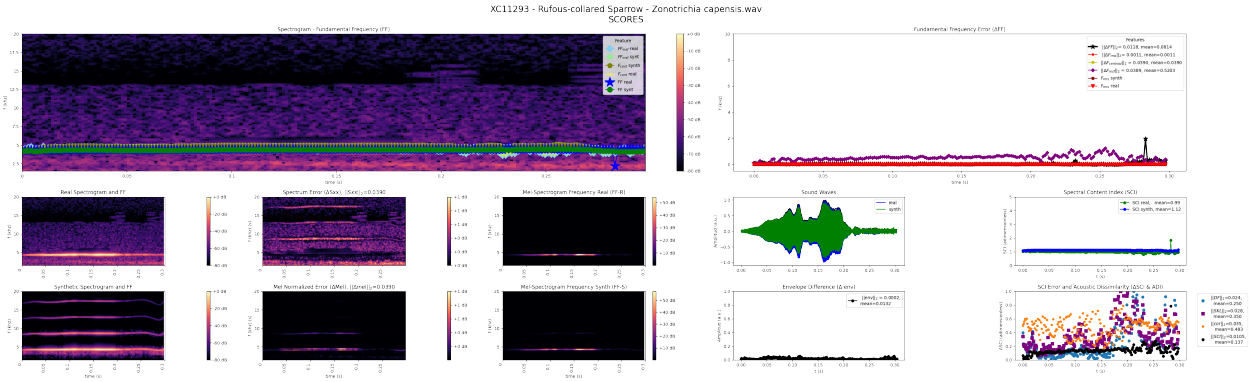


Figure 5-7: First Zonotrichia capensis syllable. A steady pitch as a pure tone around 5 kHz.

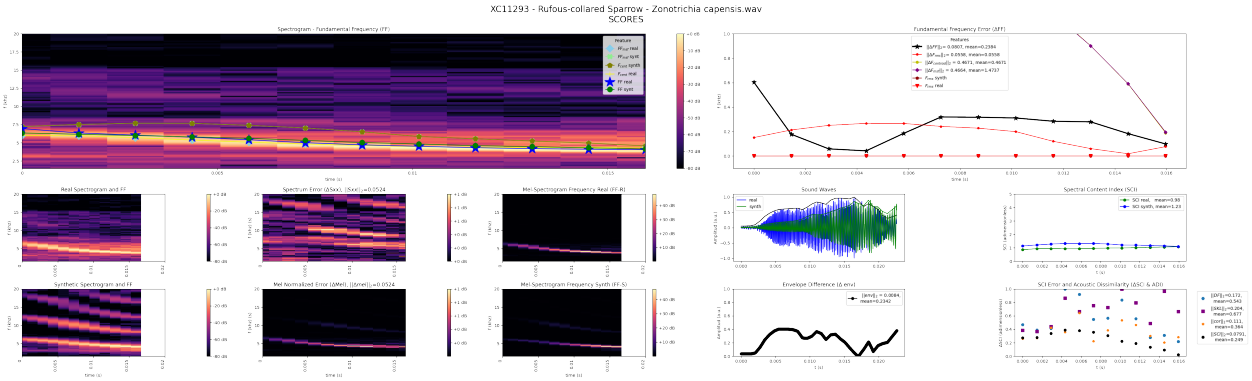


Figure 5-8: Second Zonotrichia capensis syllable. Decreasing pitch

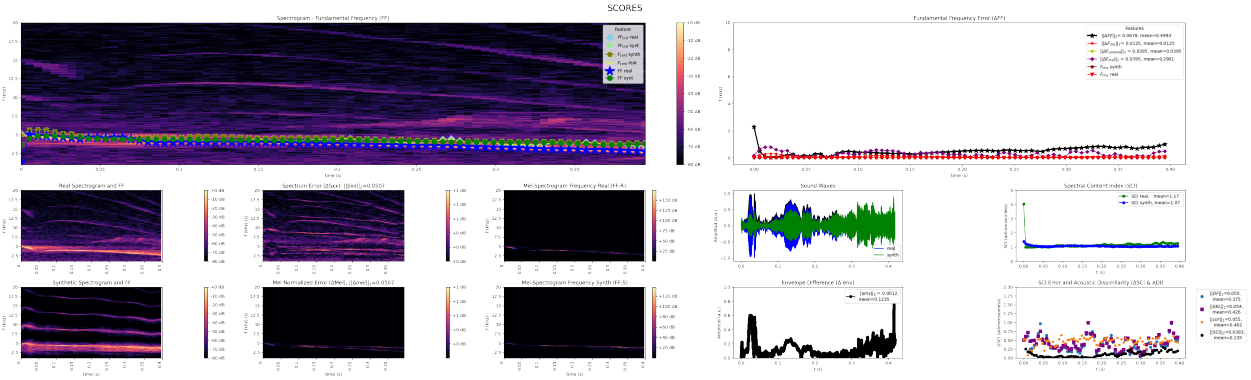


Figure 5-9: Other simple Zonotrichia capensis syllable sample

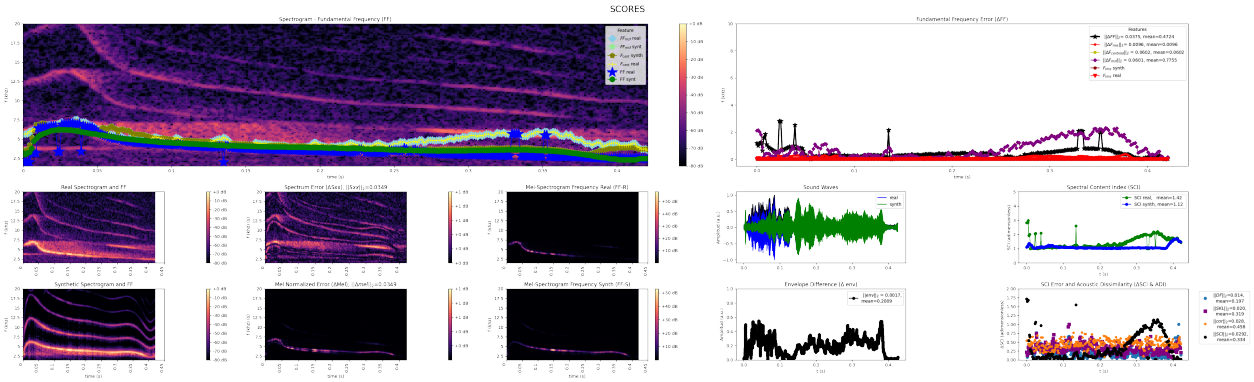


Figure 5-10: Complex Zonotrichia capensis syllable sample

5.1.3 Chunck

To analyze and generate thrilled birdsong it is necessary to do compute the STFT with small windows length[1], therefore the spectrograms have good frequency resolution but not large time resolution good enough to compute a few pitch points

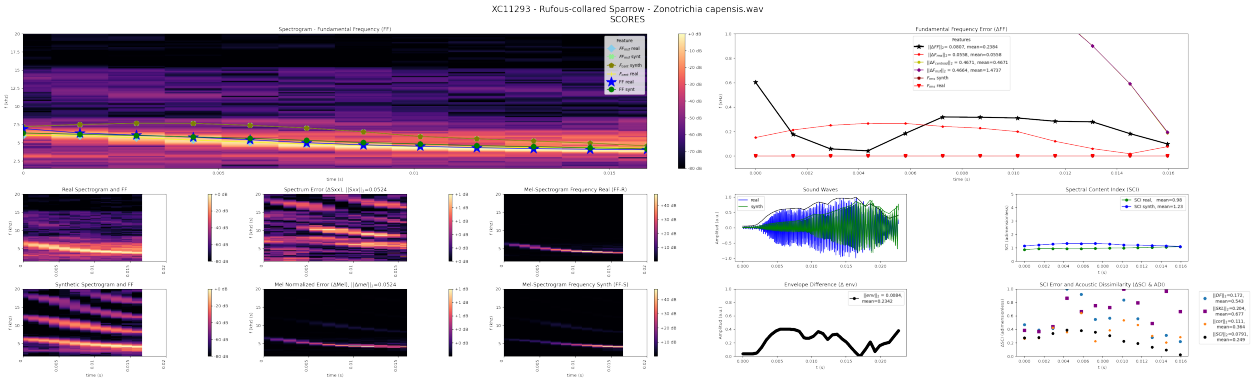


Figure 5-11: Third and small syllable of the Zonotrichia capensis sample. It is part of a series of thrilled syllables whit the same pitch shape and steady tendency

While the real spectrogram does not have strong harmonics, but has noise, the synthetic birdsong have many harmonics. This occurs because the noise is interpreted as harmonic content by the spectral content and acoustic dissimilarity indexes. The result is very good in pitch and in SCI with small errors in both quantities, other spectral variables also presented good results (f_{rms} and spectral centroid).

[1]the default Fourier window length is 1024 samples but for small time syllables a good window size can be 256 or even 128 (values below can present problems)

5.1.4 Birdsong

Computing syllable by syllable is very time consuming, since to get a good result it is necessary to adjust the pitch threshold and sometimes the samples length of the Fourier window. For this reason, the birdsong package also provides a possible synthetic birdsong generation by selecting some interval points and then using them in a birdsong method to find the optimal set of parameters values to each syllable[2]

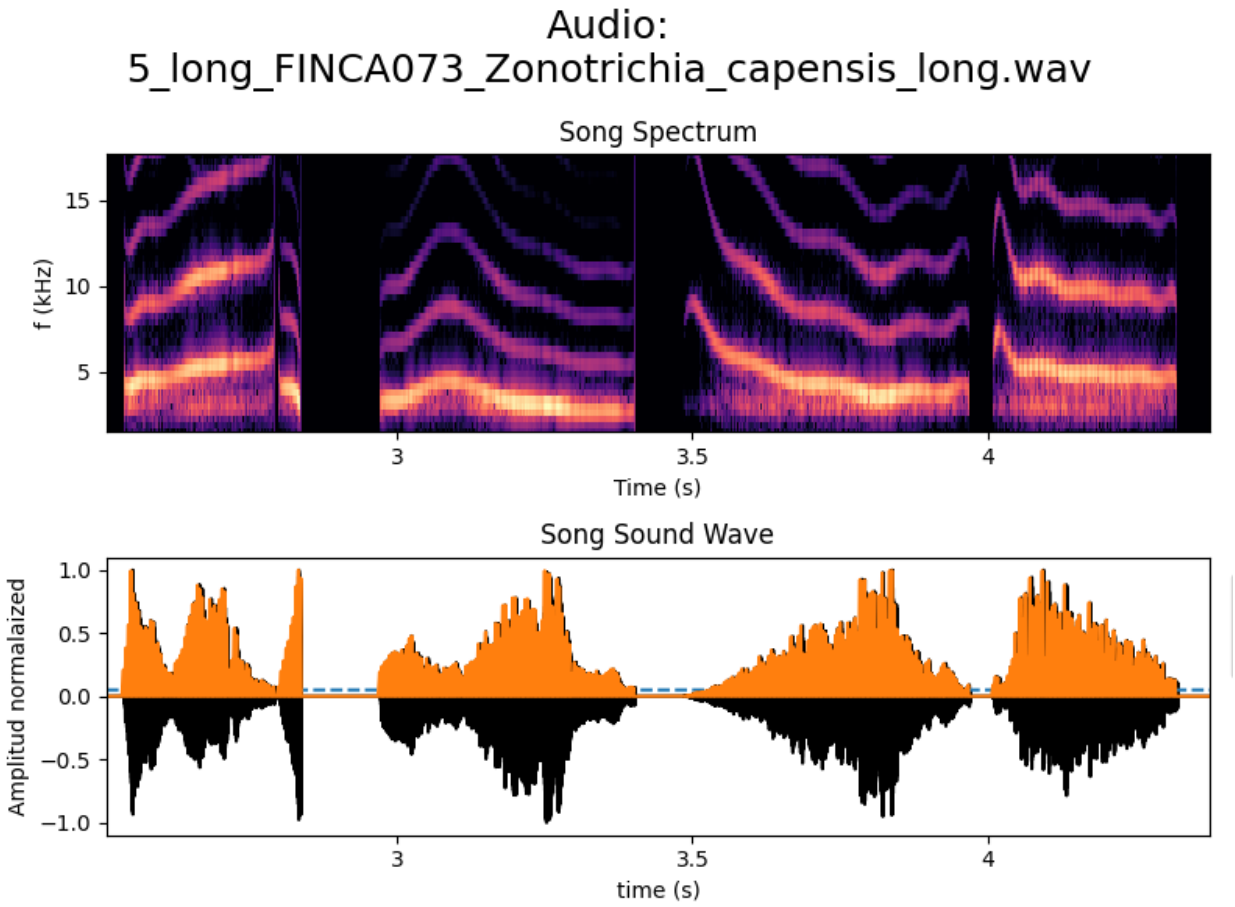


Figure 5-12: Real Zonotrichia birdsong .

[2]the optimal value of γ is the mean of all the optimal values found each syllable considering a well defined initial parameters set values, such that generates oscillations and are not too close to bifurcations

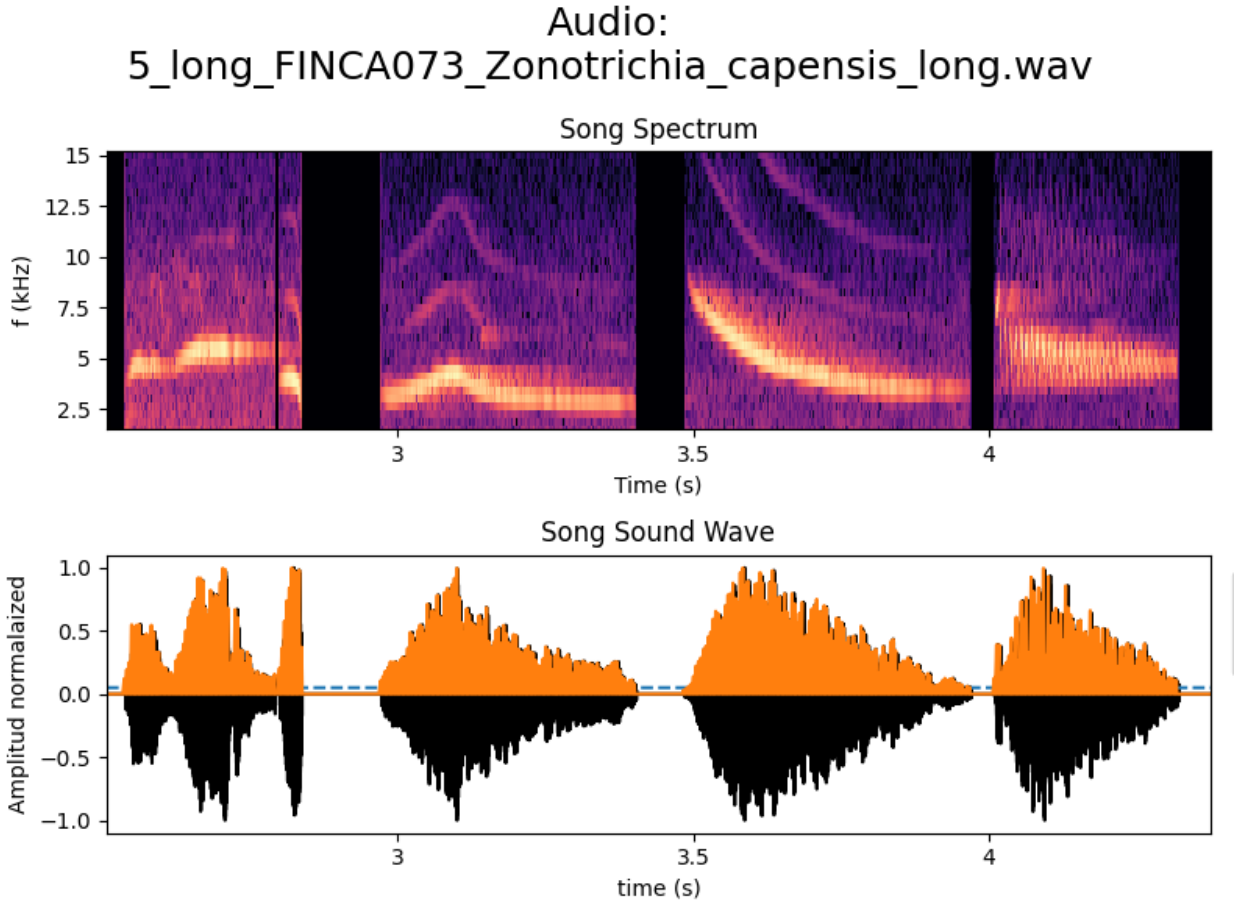


Figure 5-13: Synthetic Zonotrichia birdsong .

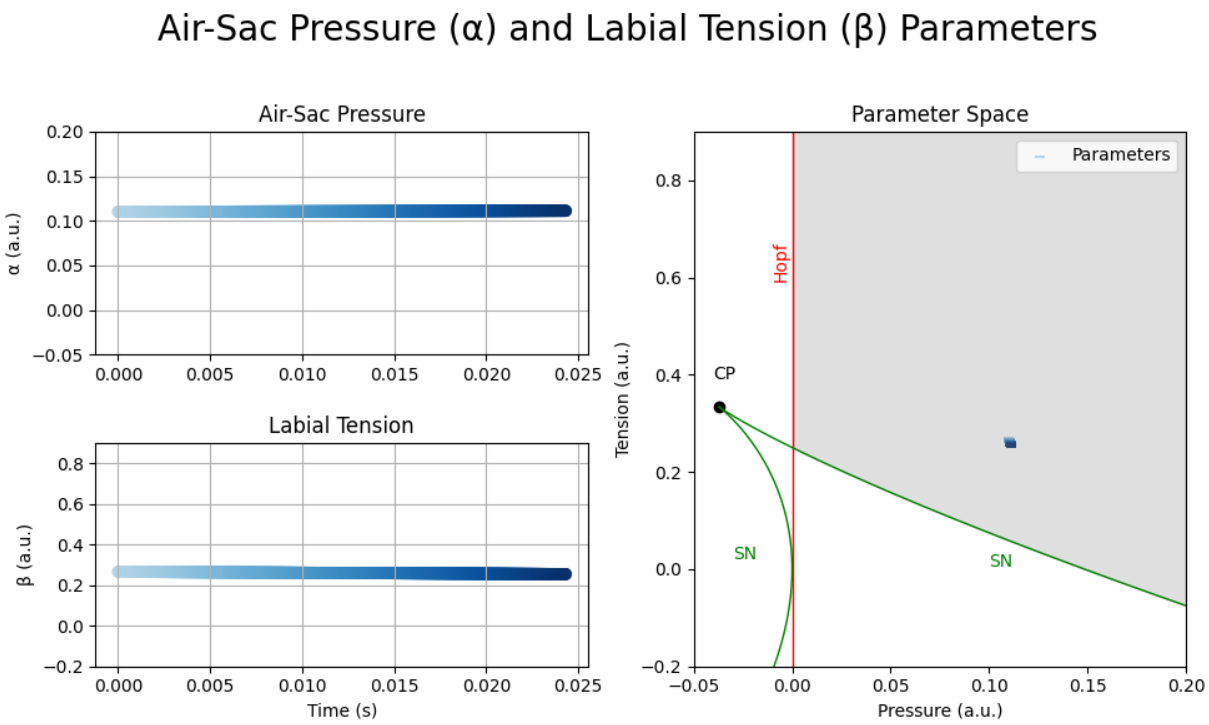


Figure 5-14: Parameters space and curve of the parameters

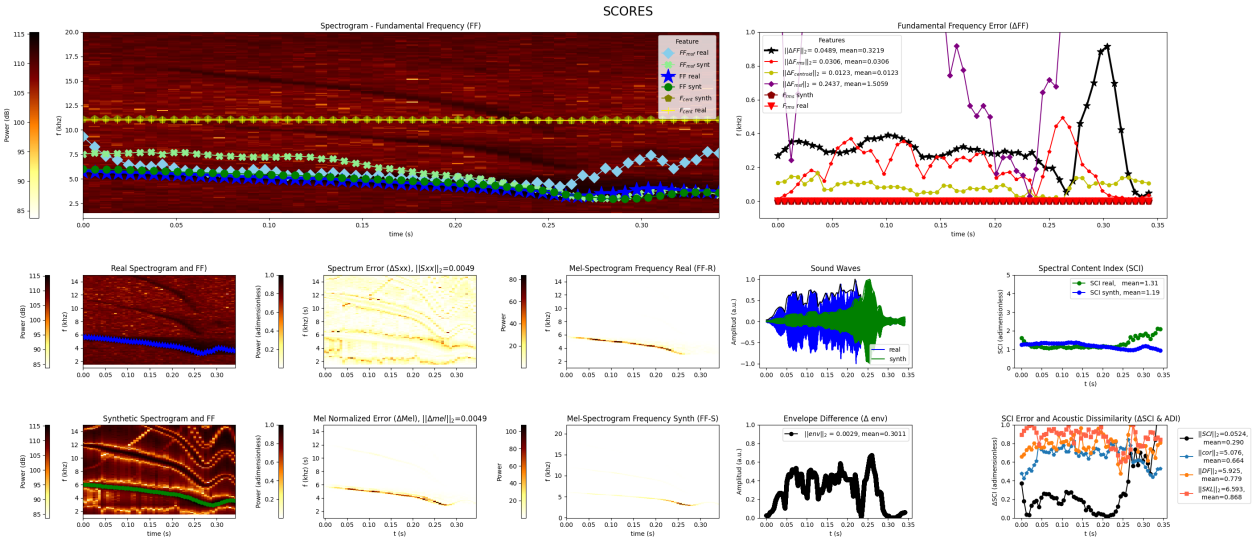


Figure 5-15: Synthetic and real syllables comparison

5.2 Evaluation

5.2.1 Consistency

1. Generate a synthetic syllable with random parameters.

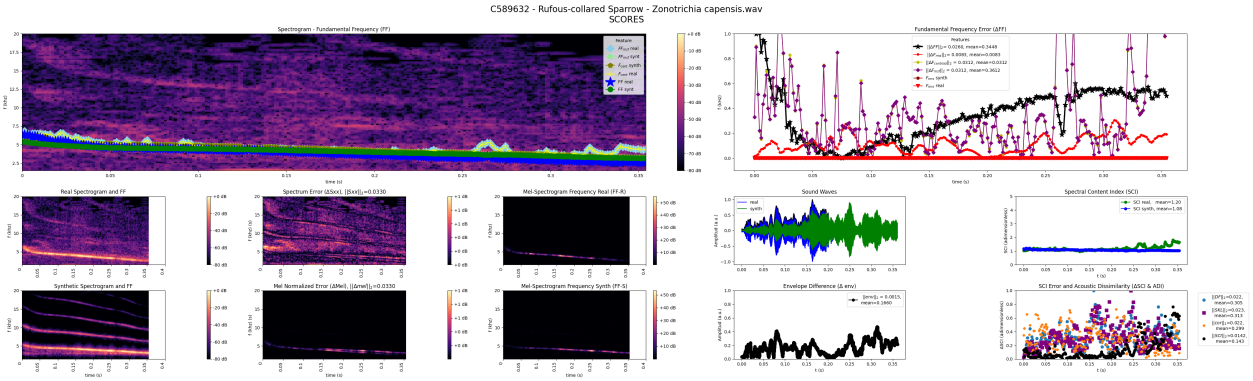


Figure 5-16: First optimal solution

2. Solve the optimization problem for the generated syllable.

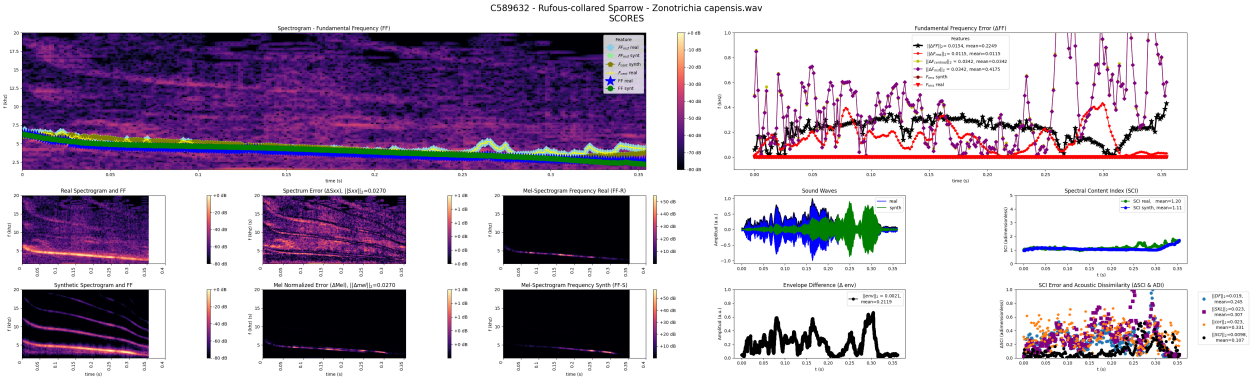


Figure 5-17: Optimal solution of the previous optimal solution

3. Compare features

   Synthetic syllables generated by the model solution twice

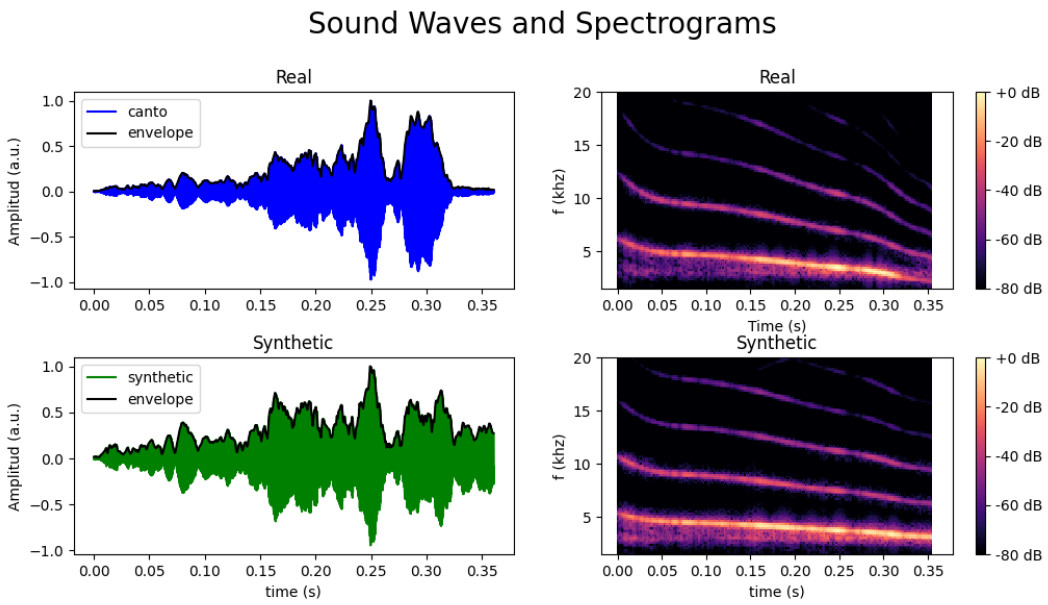


Figure 5-18: Comparison of the optimal solution (top) for the real birdsong, and the optimal solution for the synthetic syllable (bottom).

5.2.2 Uniqueness

Since the parameters space have at most 7 dimension, the optimization problem has not a single optimal and different parameters sets may generate the same syllable. The proposed method reach a local minima by solving the optimization subproblems.

5.2.3 Generalization

For the Zonotricha capensis the birdsong package works efficiently and generated good qualitative and quantitative results, comparable birdsongs. To testing the model generalization other species are used, they are listed in order of pitch complexity from the easiest one.

1. Tapaculos (Rhinocryptidae): pitches are elementary functions that can be approximate as linear or quadratic functions

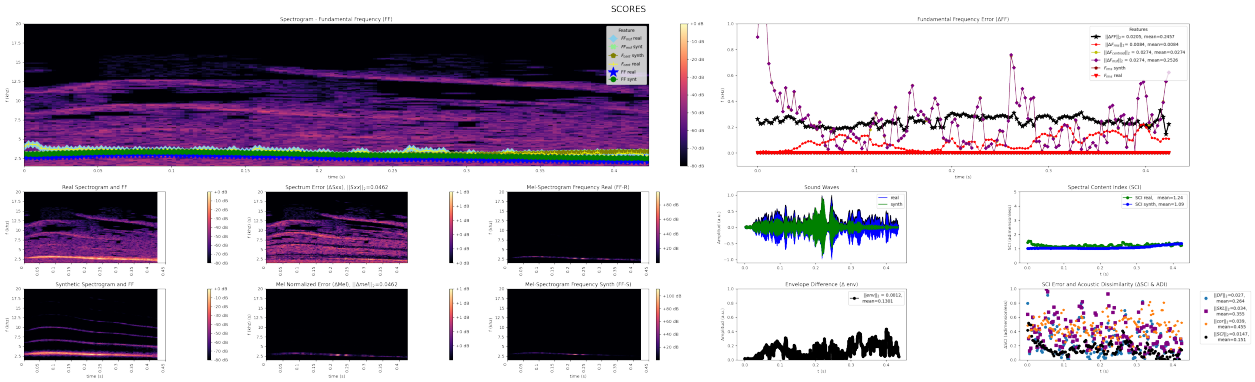


Figure 5-19: Simple Tapaculos syllable sample

2. Euphonia laniirostris

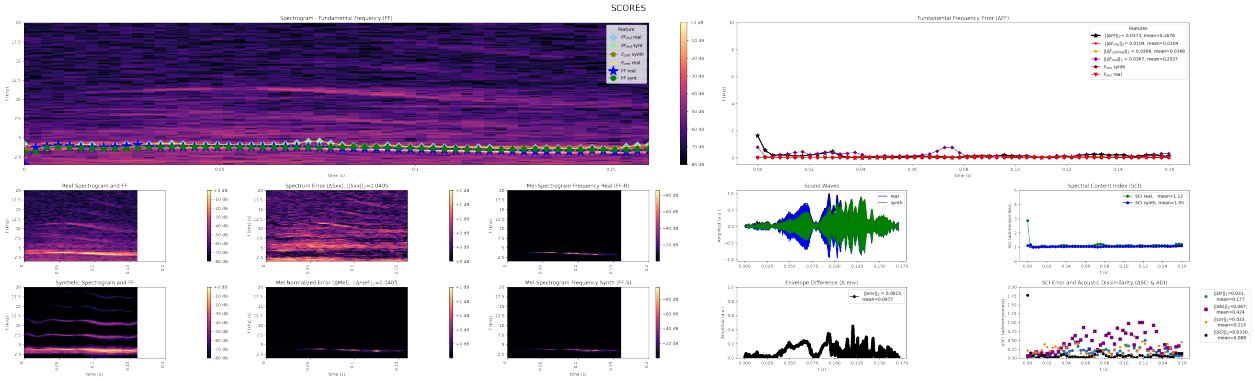


Figure 5-20: Simple Euphonia laniirostris syllable sample

3. Mimus gilvus

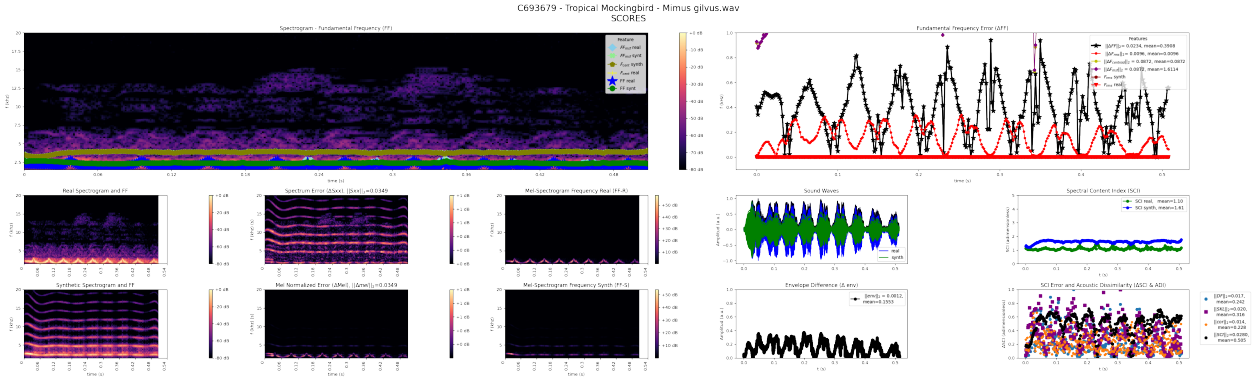


Figure 5-21: Simple tropical Mockingbird - Mimus gilvus syllable sample

The audios are store on xeno-canto audios collection, you can fin the used audios at the collection dissertation.

6 Conclusions

6.1 Conclusions

- The model successfully simulated several syllables of Zonotrichia capensis with different sound quality. The best sounds to generate are the longer, simpler and clear syllables which were simulated with high accuracy. The thrilled syllables can be well-generated using chuncks, small parts of syllables, but it requires tuning the pitch threshold.

- The most problematic and difficult syllables are the noisy and with high spectral content audios, in which strong harmonics are present making the pitch computing hard or even impossible to compute correctly. Although for some audios is sufficient to change the pitch threshold detector, it does not work for all of them.

- The SCI score gives comparable results to finding the optimal pressure parameters coefficients, however it is not always sufficient since the noise can be interpreted as harmonics or spectral content. An improvement is to refine the objective function that find these parametric coefficients.

6.2 Boundaries

- Since the model implemented in this work depends on pitch extraction and fundamental frequency computation, its accuracy is limited. In fact, if the fundamental frequency is computed incorrectly, the proposed model will not find an optimal solution because the score function is not correctly defined.

- Although there are many available birdsong audios on the web, not all of them have clear birdsongs. The best birdsong audios for the model are the audios with great sound quality, clear samples.

6.3 Future Works

- Explore more species and classify them by their parameters values, air sac pressure and tension labia. It can provide a new novel way to classify birds by their syrinx parameters.

- Creating a more realistic model with two pairs of lateral labiums, two syrinx, will allow to create more complex syllables in which two different pitches are present at the same time. How to extract two pitches?
- Use advanced optimization algorithms that involve the calculation of the Jacobian and Hessian of the objective function (that depends of the bifurcation curves). Chuncks can be used to linearize the system locally and make the problem well behaved, may imply a better convergence rate.
- Search and explore physical models of production of sound for other type of animals vocalization: frogs, insects, fishes, etc, or even vocal folds.
- Improve package syllables detection and spectral features computation, implement an algorithm to compute complex fundamental frequencies and its harmonics. Parallelize objects, make them pickleable.

UNIVERSIDAD
NACIONAL
DE COLOMBIA