Processamento e identificação automática de cantos de pássaros

Felipe Felix (f^2)

May 22, 2017

IME-USP

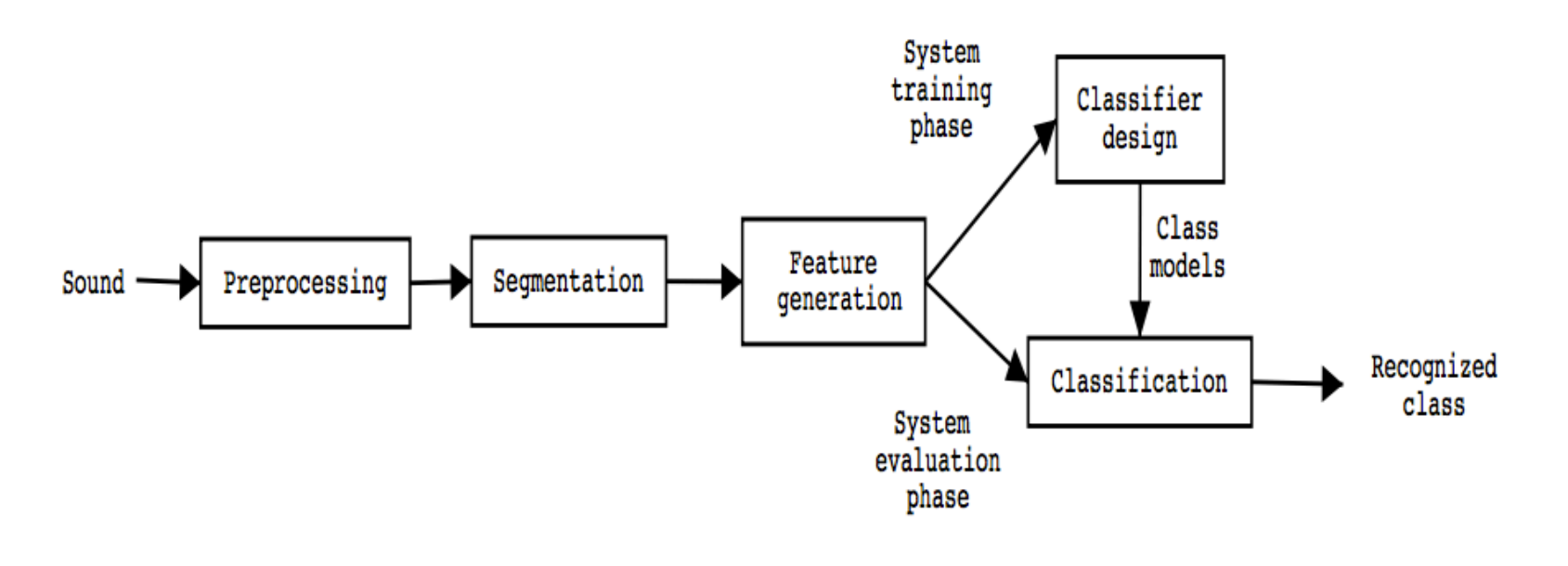
Table of contents

Introduction

Introduction

- Automatic identification and classification system for bird species using their sounds.
- Estimation of the range, population size, and population trends using ARUs.
- Biodiversity indicators.

Audio classification general structure



Introduction - Bird Sounds

- Usually divided in three categories: *call*, *song*, *mechanical sounds*.
- *Calls* usually refer to simple frequency patterns of short monosyllabic sounds.
- *Songs* are longer than calls, acoustically more complex, and often have a modular structure.
- Hierarchical levels of bird song: phrases, syllables, and elements (or notes).
- *Element* is the smallest separable element in spectrogram.
- *Syllables* are produced by one or more elements.
- Series of *syllables* that occur together in a particular pattern is called a *phrase*.

Example

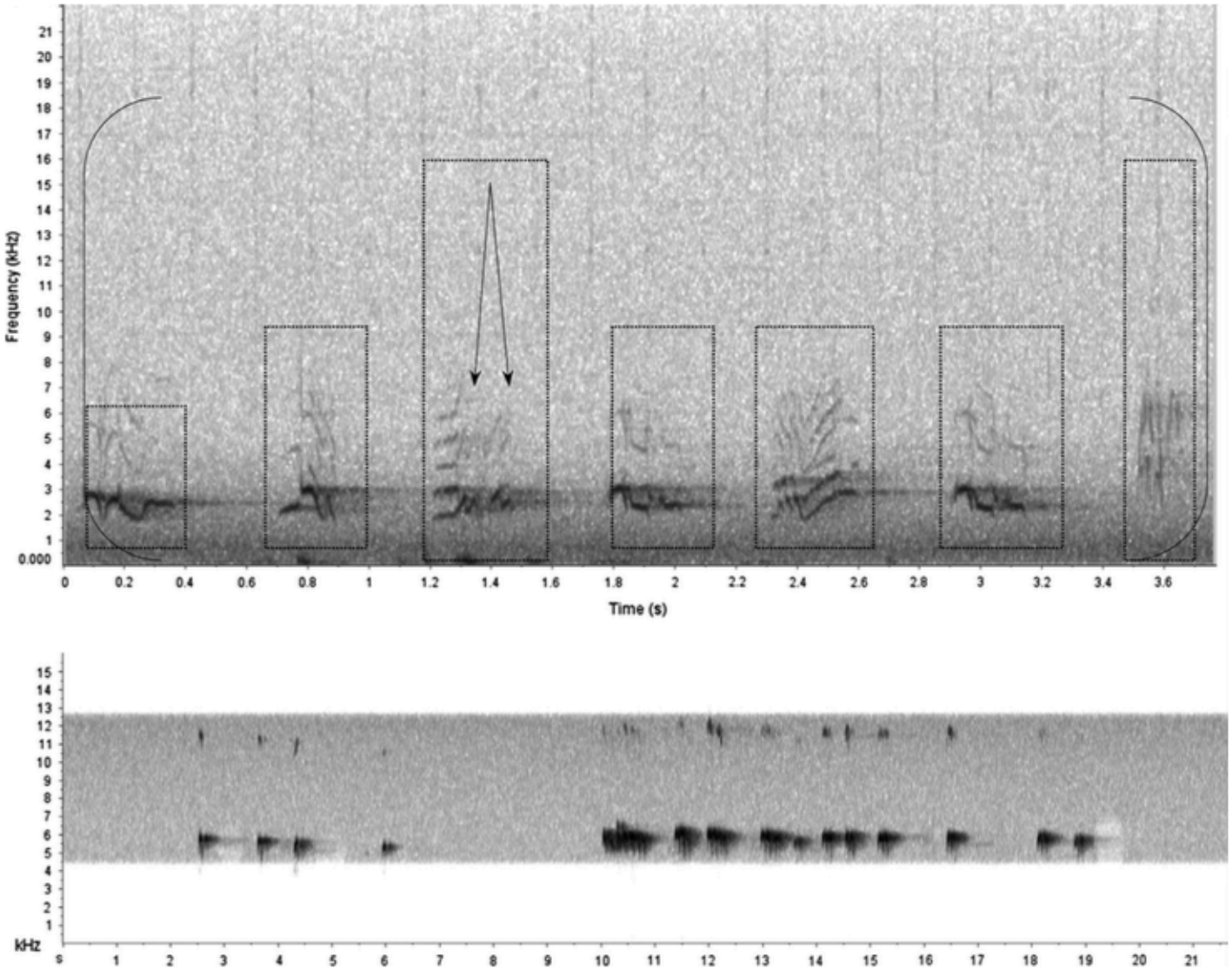


Fig. 1. Spectrograms of target species. Top row: song of American Robin – *Turdus migratorius*, elements (arrows), syllables (boxes), song (parenthesis). Second row: Kingfisher calls.

Example

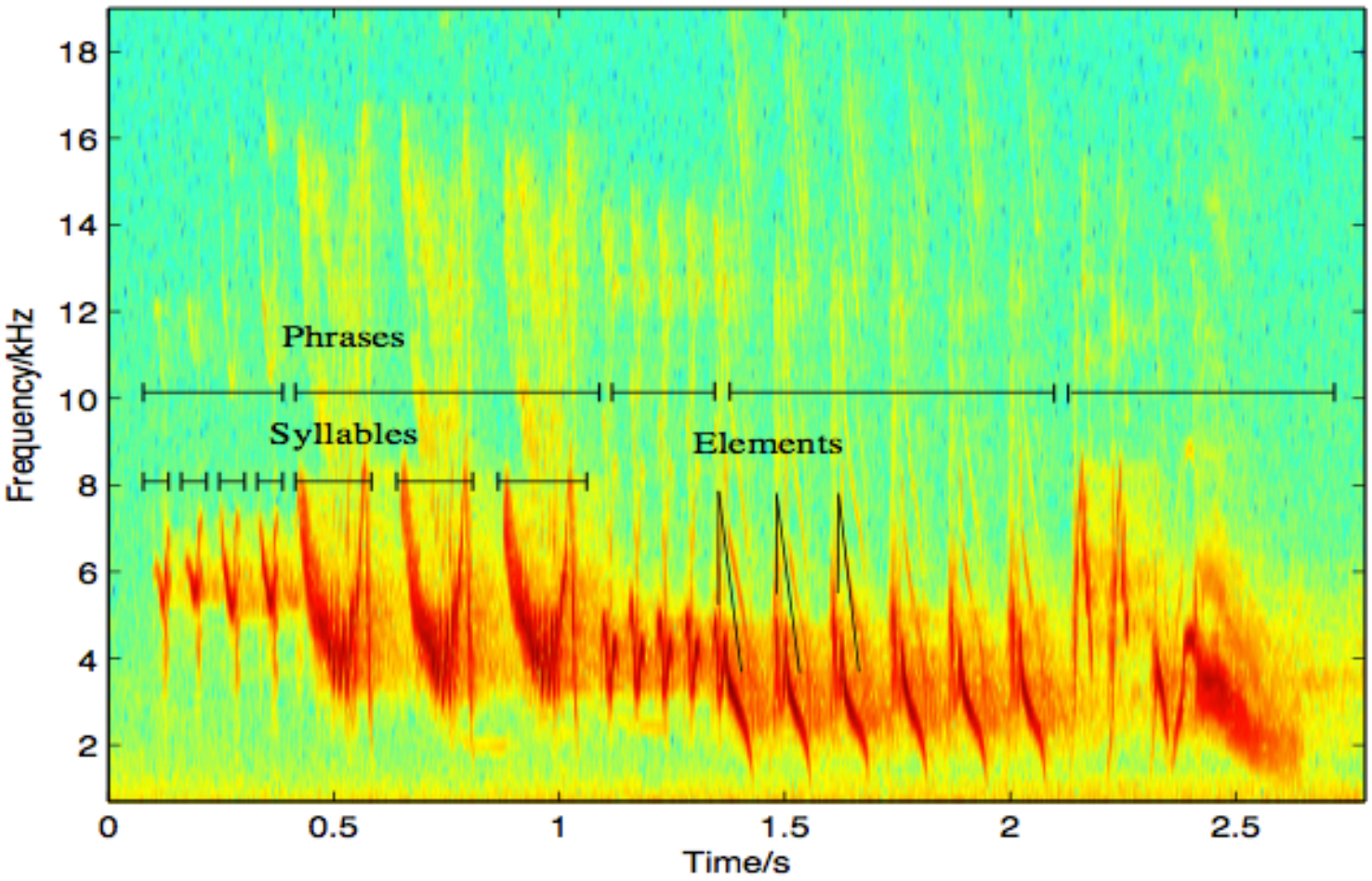


Figure 2.5: Hierarchical levels of common chaffinch (*Fringilla coelebs*) song. The y-axis represents frequency in Hz and x-axis time in seconds.

Segmentation Algorithm

Algorithm

- Based on the energy of audio file and the noise level estimate.
- **First phase**: onset and offset of the syllable candidates are calculated.
- Near onset and offset areas of syllable candidates energy envelope curve commonly fluctuates around the threshold, which causes short erroneous candidates (*border effect*).
- **Second phase**: syllable candidates enough close to each other in temporal domain are connected to a single syllable.
- **Last phase**: true syllable detection; where too short candidates are omitted.

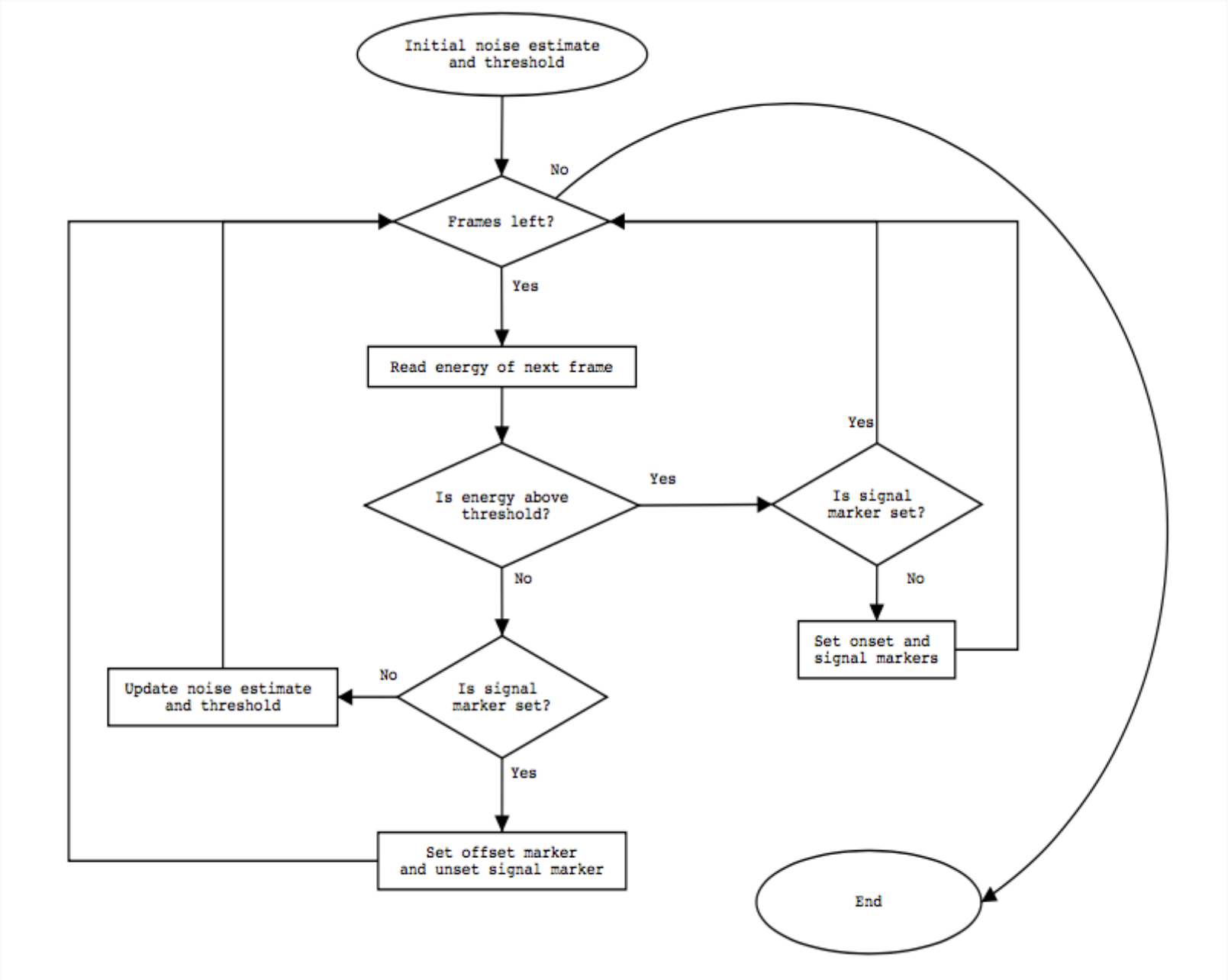
Onset/Offset detection

- File is divided into the overlapping frames.
- Frames are windowed using the hanning window.
- The energy envelope $E(m)$ of the signal $x(n)$ in decibel scale is calculated as

$$E(m) = \sum_{i=1}^{N} 20log_{10}|x_i[n]|^2 \tag{1}$$

where $x_i[n]$ is i:th frame and N is the total number of frames of the signal $x(n)$ (?).

Flow diagram of syllable search algorithm



Example

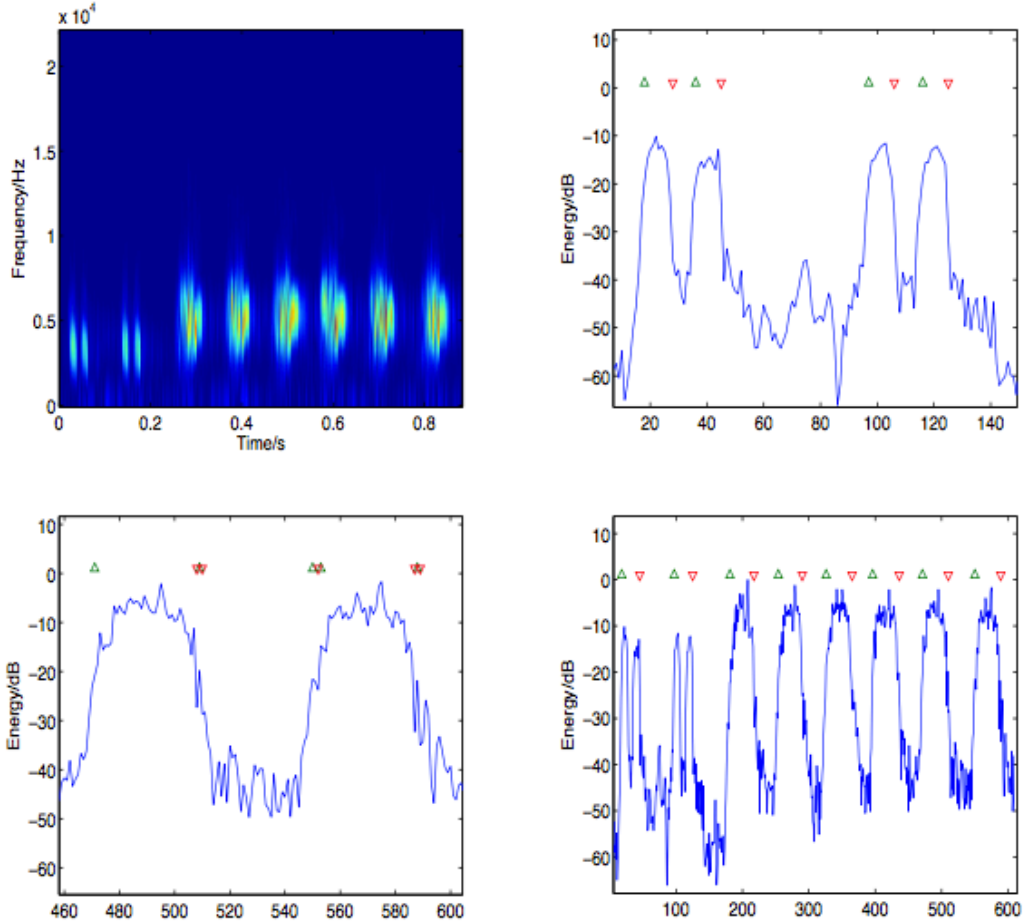


Figure 3.3: Segmentation of bird song using short time signal energy method. Top left panel shows spectrogram of eight syllables of a song of Sedge Warbler (*Acrocephalus Schoenobaenus*). Top right and bottom left panels shows segmentation result before merging syllable candidates phase. Panels illustrates respectively syllables consist of two pulses and the border effect in two pulses. Final segmentation result is presented in the bottom right panel.

References

[1] Potamitis, Ilyas, et al. ”Automatic bird sound detection in long real-field recordings: Applications and tools.” Applied Acoustics 80 (2014): 1-9.
[2] Fagerlund, Seppo. Automatic recognition of bird species by their sounds. Diss. Helsinki University of technology, 2004.