香港中文大學（深圳）
The Chinese University of Hong Kong, Shenzhen

EIE3510 Digital Signal Processing-Term Project

Composition and Instrumental Music Generation from Bird Songs

Yang Jiao 118010121 Jifei Zhao 118010437

13 December 2021

Introduction

Motivation

In the AI generation, researchers started exploring the boundary between machine intelligence and artworks by using the machine to generate music. Previous studies mainly focused on deep learning methods, which need massive data.

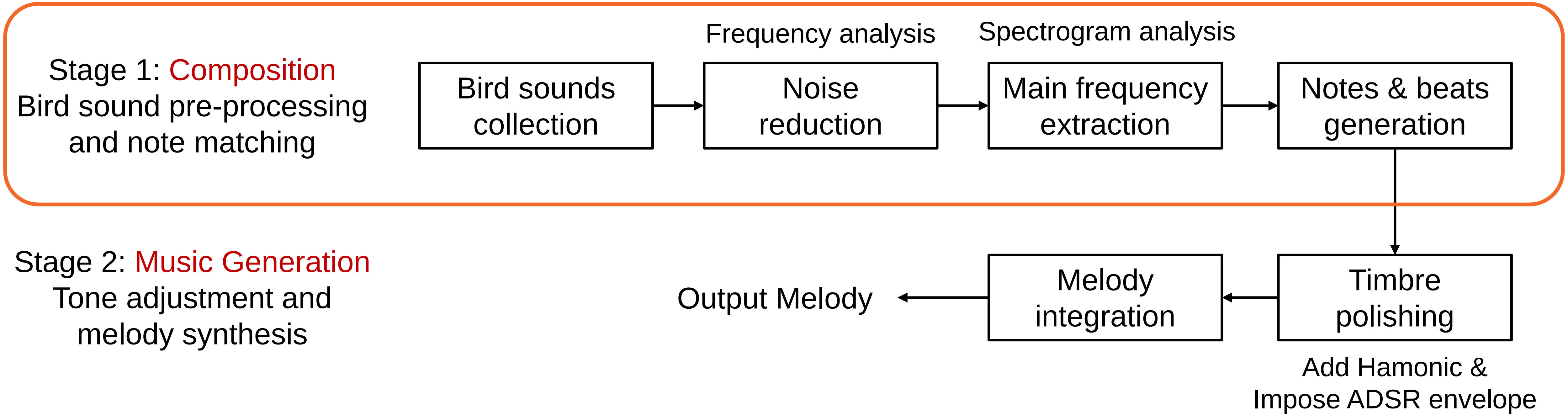
- We hope to design an explicit and lightweight model with only a tiny amount of unmusical audio input.
- Inspired by nature, we investigate the similarity between the singing of birds and instrumental music.

Goal

1. Map the bird sounds to a series of notes (Composition)
2. Play the melody by mimicking the timbre(音色) of piano (Music Generation)

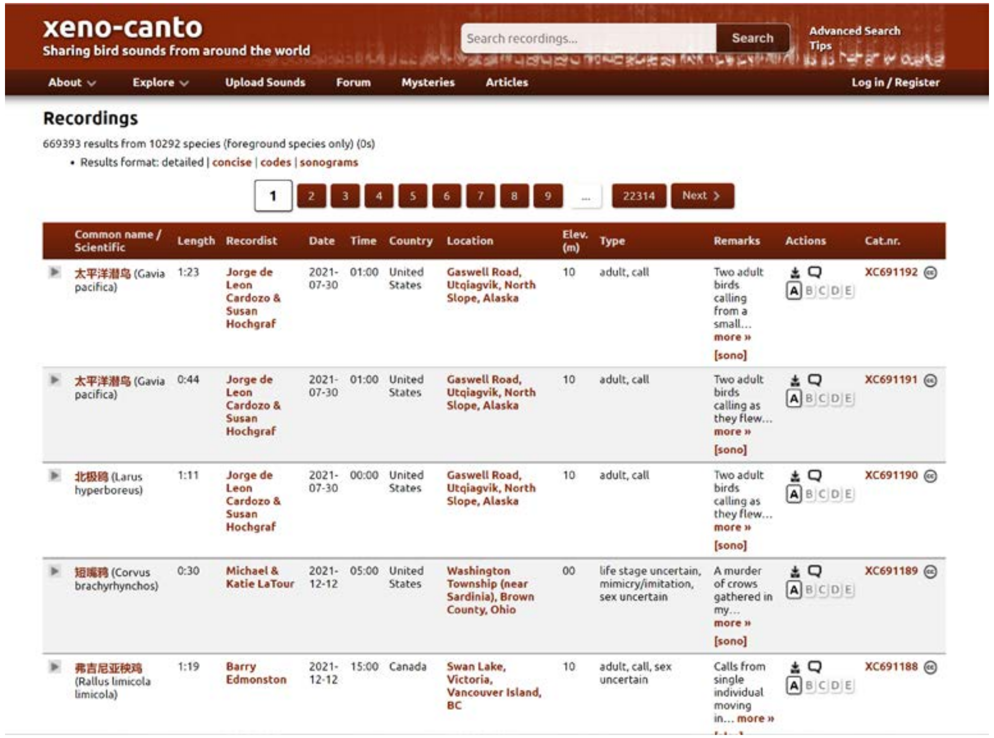
Methodology

Framework of whole project :

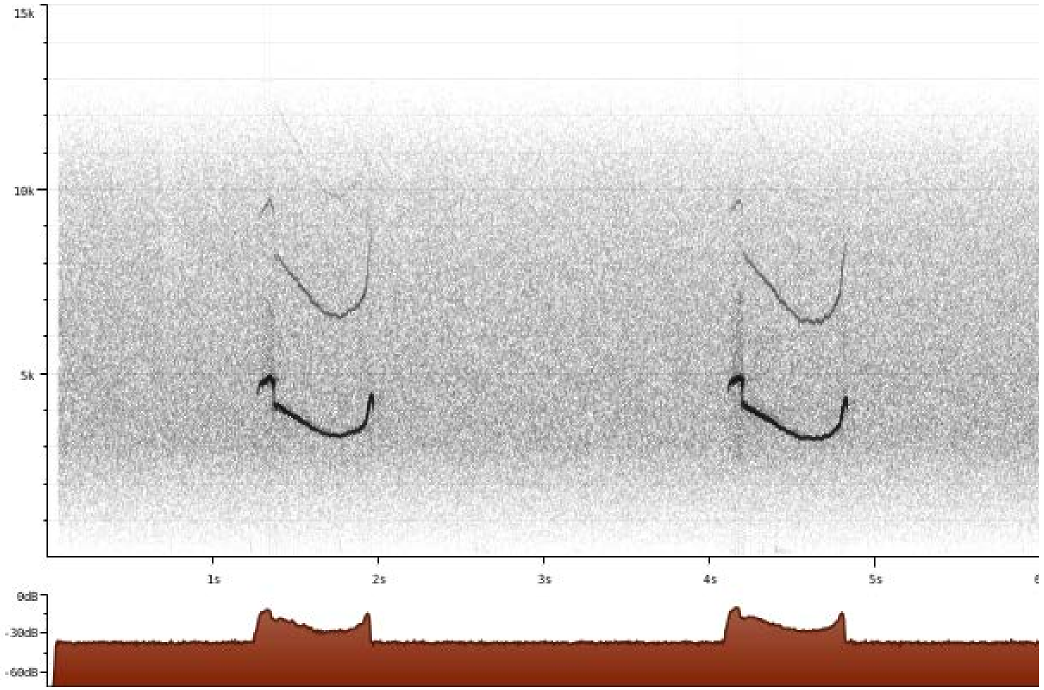


Stage 1—Bird sounds collection

Dataset :



https://xeno-canto.org/



Spectrogram information

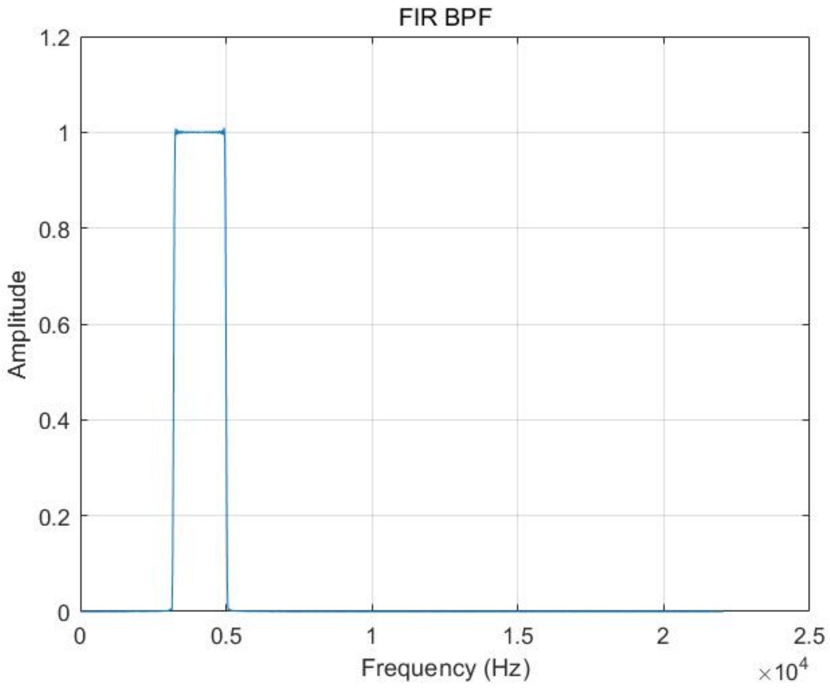
Stage 1 — Noise Reduction

Ideal band-pass filter

- Perfect square wave
- Cutoff frequency: fc1 = 3200 Hz, fc2 =5000 Hz

FIR band-pass filter

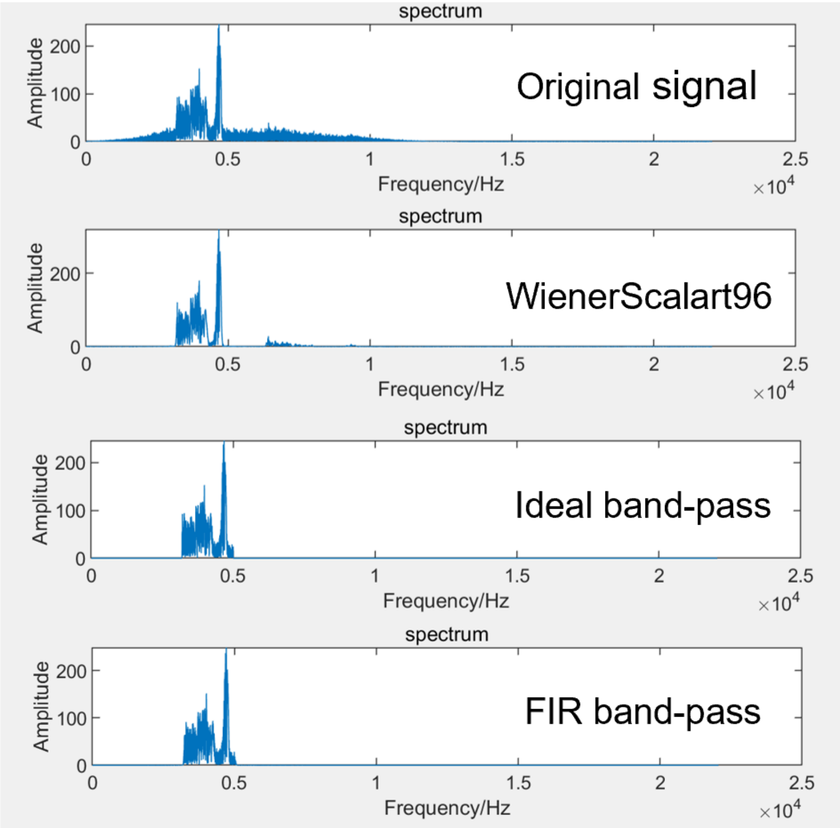
- Width of transition band: 100 Hz
- Ripples: [0.01, 0.01, 0.01]
- Cutoff frequency: fc1 = 3200 Hz, fc2 =5000 Hz
- Kaiser window with M = 986, β = 3.3953



Stage 1 — Noise Reduction

Comparison between filters

- **Filtering effect in frequency domain:**



- **Filtering effect on the basis of SNR and time complexity:**

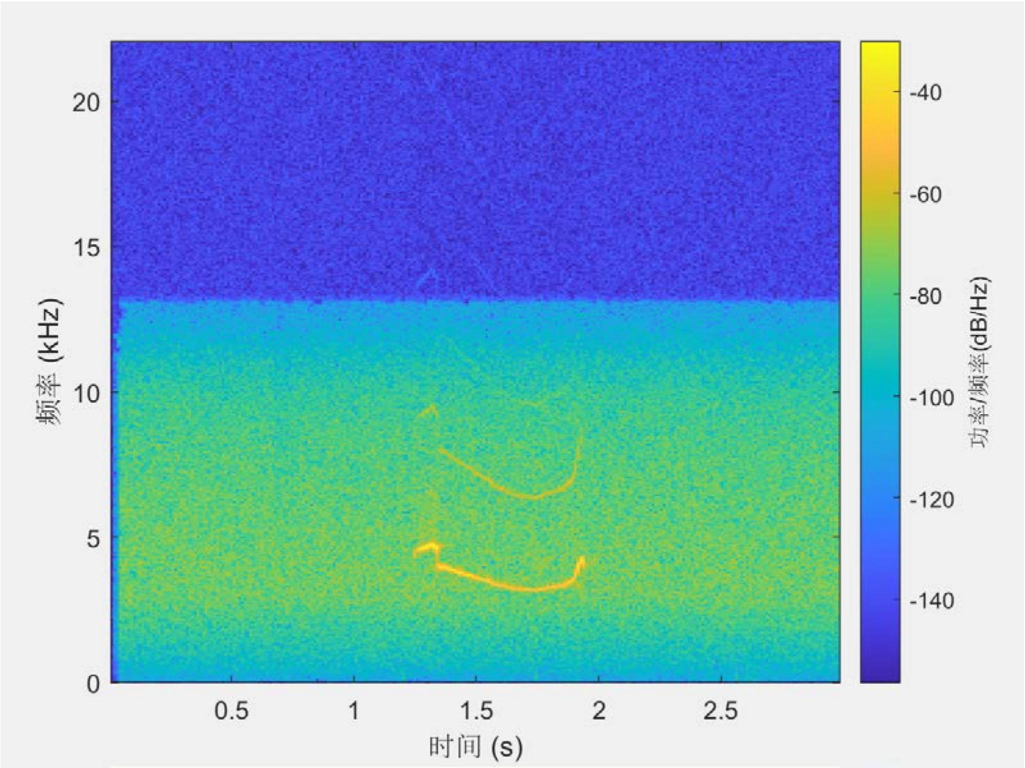
Filter	SNR (dB)	Time complexity
Ideal band-pass	10.8853	$\mathcal{O}(N \log_2 N)$
FIR band-pass	10.8257	$\mathcal{O}(N)$
[P. Scalart, 1996] Wiener estimation	13.2802	/

Stage 1 — Main frequency extraction

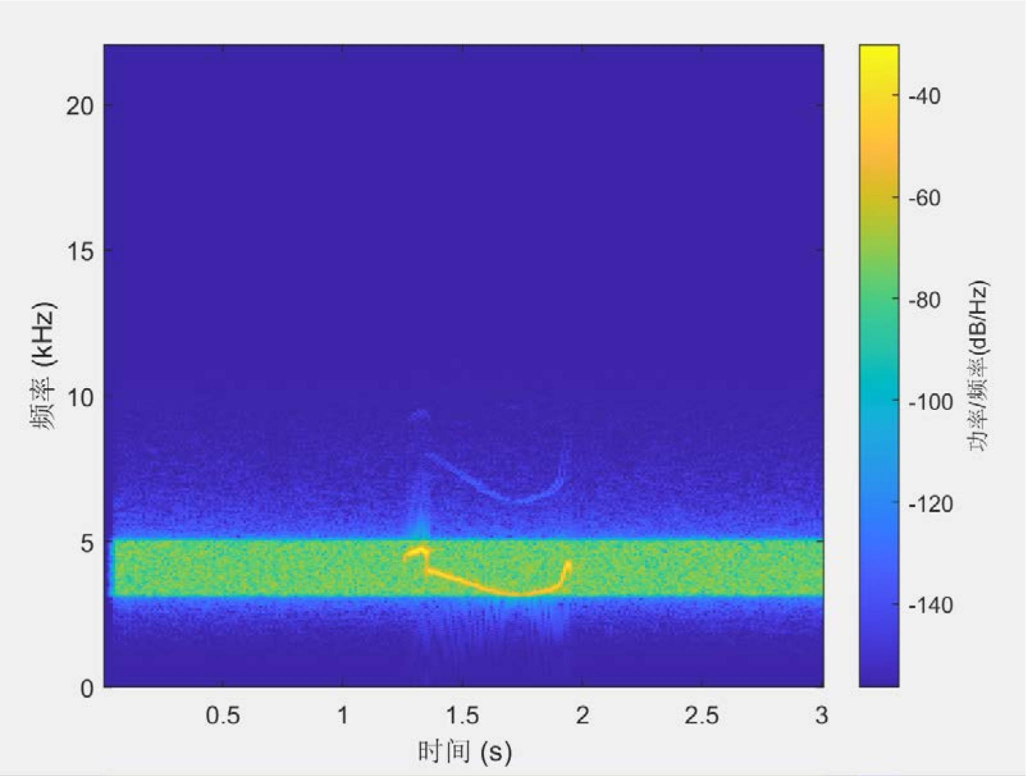
Spectrogram analysis

We apply Short-time Fourier transform with:

- window length L= 1000 (23 ms Hamming window),
- window moves by R= 500 samples each time.



Original Spectrogram



Spectrogram after FIR BPF

Stage 1 — Notes & beats generation

Notes

- Starting from a frequency on the idealized standard piano tuned to A_4 (440 Hz)
- Key number can be calculated from

$$Key(n) = \mathrm{round}(12\log_2\left(\frac{f_{main}(n)}{440\ \mathrm{Hz}}\right)) + 49\ \mathbf{-8}$$

The 49th key on piano

Key number	MIDI note	Helmholtz name[5]	Scientific name[5]	Frequency (Hz) (Equal temperament) [6]
79	99	d♯′′′′/e♭′′′′	$D\sharp_7/E\flat_7$	2489.016
78	98	d′′′′	D_7	2349.318
77	97	c♯′′′′/d♭′′′′	$C\sharp_7/D\flat_7$	2217.461
76	96	c′′′′ 4-line octave	C_7 Double high C	2093.005
75	95	b′′′	B_6	1975.533
74	94	a♯′′′/b♭′′′	$A\sharp_6/B\flat_6$	1864.655
73	93	a′′′	A_6	**1760.000**

Decrease the frequency by an octave

Stage 1 — Notes & beats generation

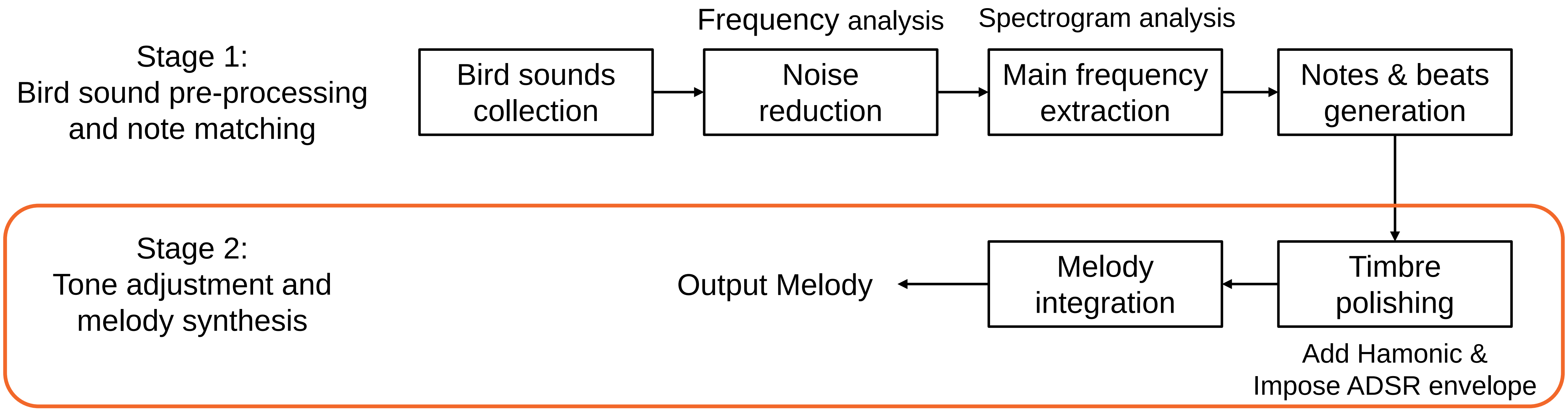
Beats

- Whole note duration = 1s
- Note usually divided by a factor of powers of two
- Whole note, half note, quarter note, and eighth note are considered
- Mapping rule between number of repetitive frequency levels & standard duration of notes

Number of repetitive frequency levels	Note value	Time duration (s)
1	Eighth note	0.125
2	Quarter note	0.25
3, 4, 5	Half note	0.5
6, 7, 8, 9, 10	Whole note	1

Stage 2

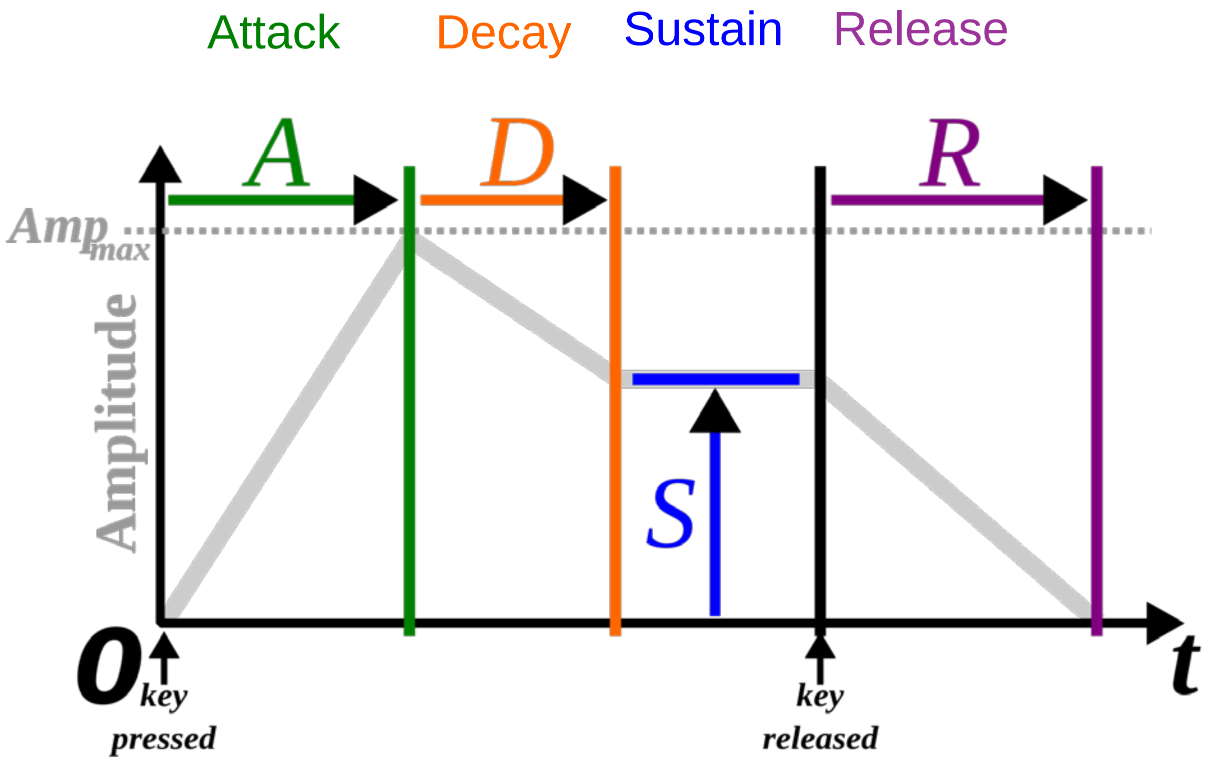
Tone adjustment and melody synthesis



Stage 2 — Timbre Polishing

How is timbre determined for a kind of instrument?

- Envelop: ASDR model
- Superposition of high order harmonics



Stage 2 — Timbre Polishing

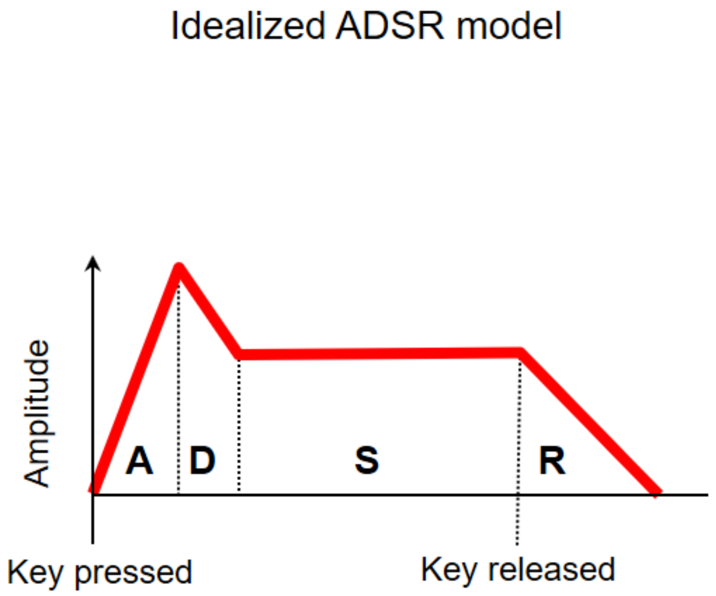
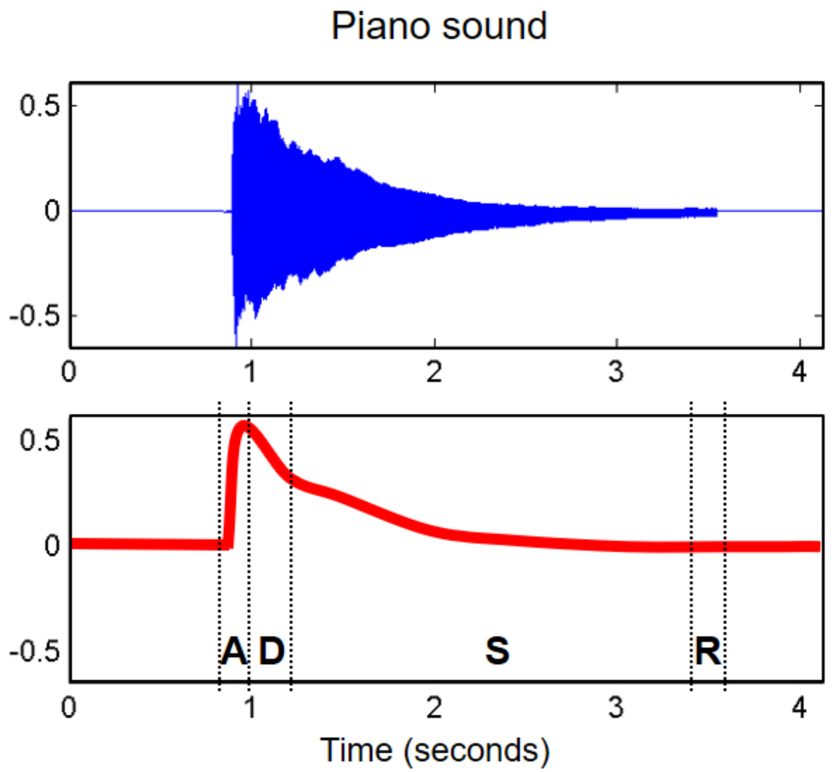
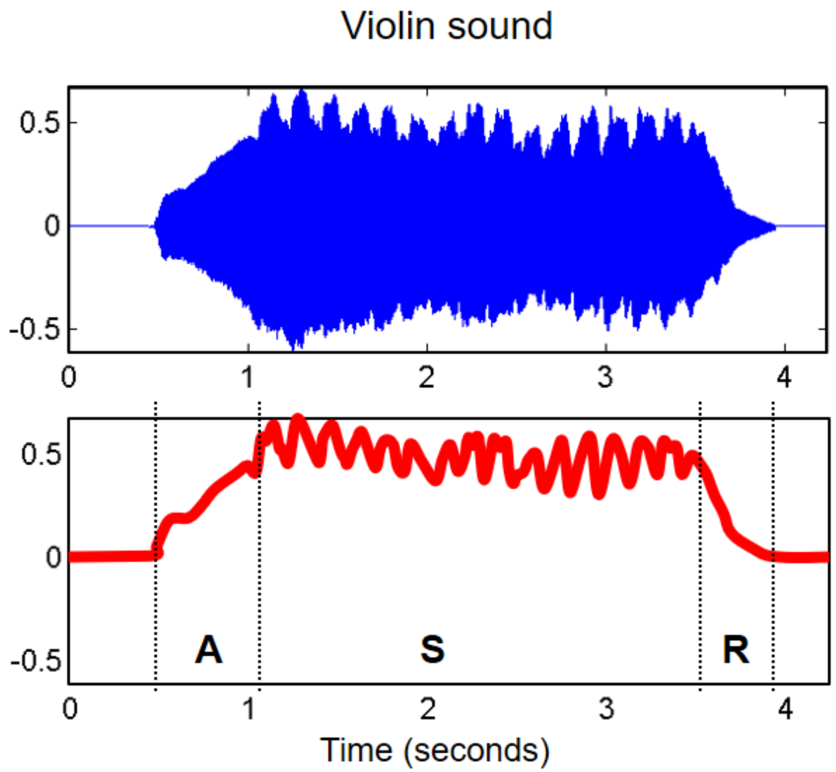


Figure 1.22b and Figure 1.23
from [Müller, FMP, Springer 2015]





- Very short attack time
- Decay drastically
- Envelop approximated by an exponential function

- Long attack time for set up
- Long sustain with vibration and tremolo
- Difficult to mimic by an elementary function

Stage 2 — Timbre Polishing

Different envelop shapes

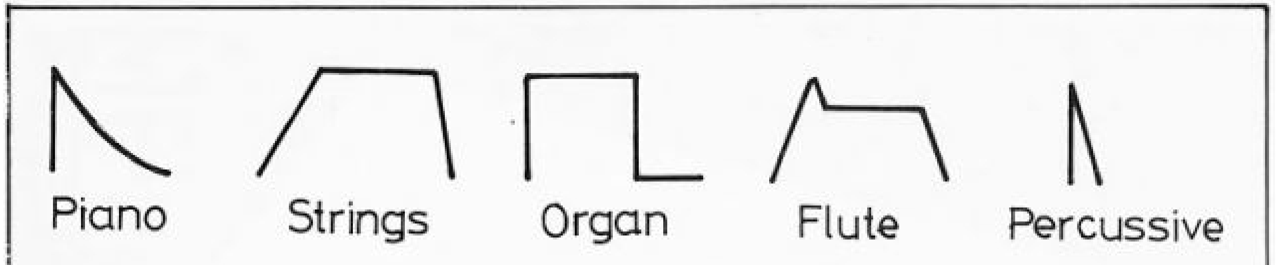
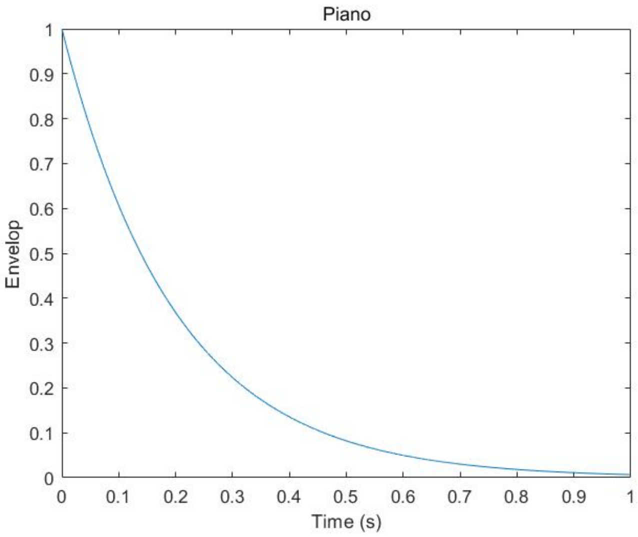
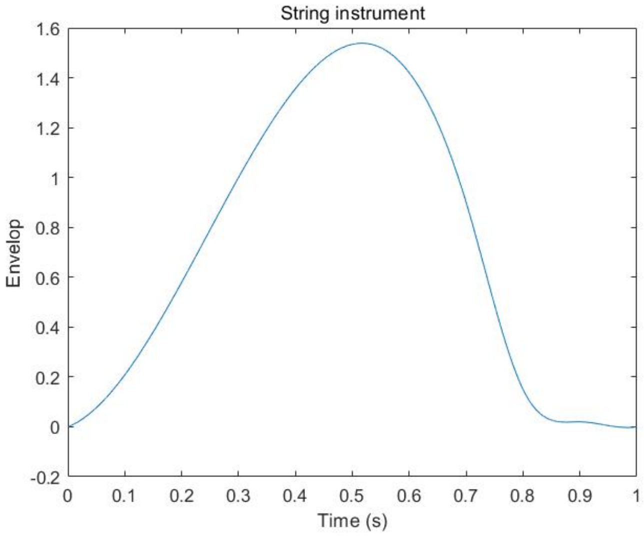


Figure 5. The Steve Howell Guide to Instruments and their Envelope Shapes.

Piano



Violin




```
L = 1;
y = [0, 1, 0.9, 0.15, 0.02, 0.005, 0];
x = [0, 0.3, 0.7, 0.8, 0.9, 0.95, 1];
fs = 44100; Ts = 1 / fs;
t = 0:Ts:L;
env = spline(x*L, y, t);
```

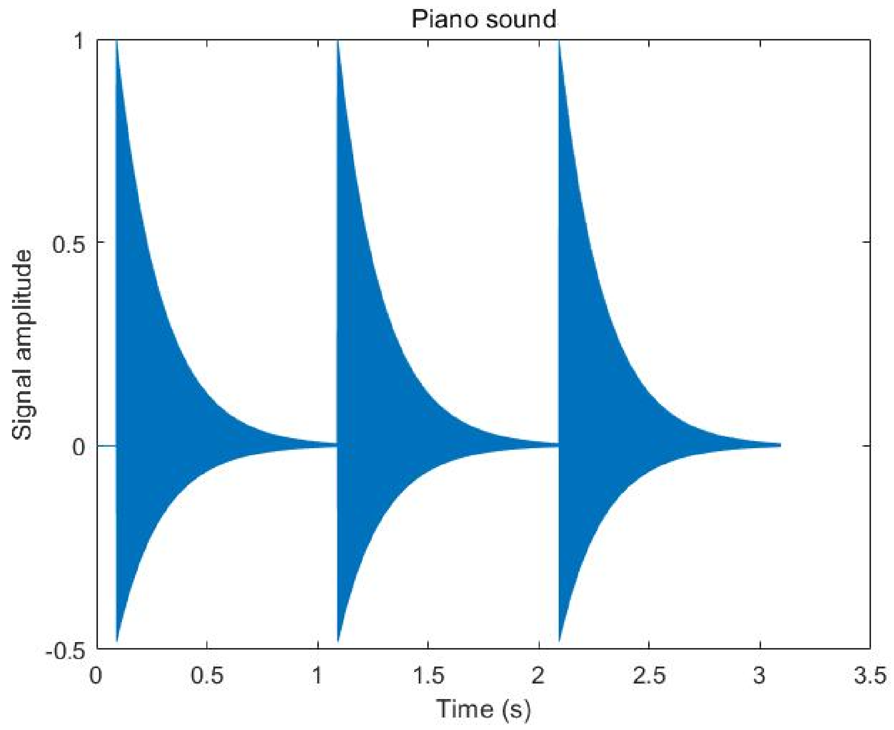
MATLAB code to generate the envelop of violin

Stage 2 — Timbre Polishing

High order harmonics

- Added by empirical tuning
- Just an approximate model. Professional music synthesis may sample and analyze the real instrumental sound first

- At fundamental frequency (main frequency) f_c, the signal for one note can be expressed as:



$$y = \underbrace{env(t)}_{\text{Envelop function}} \times \underbrace{\left(\cos(2\pi f_c t) + 0.4\cos\left(4\pi f_c t + \frac{2\pi}{3}\right) + 0.15\cos\left(6\pi f_c t + \frac{4\pi}{3}\right)\right)}_{\text{sinusoidal signal at } f_c \text{ with high order harmonics}}$$

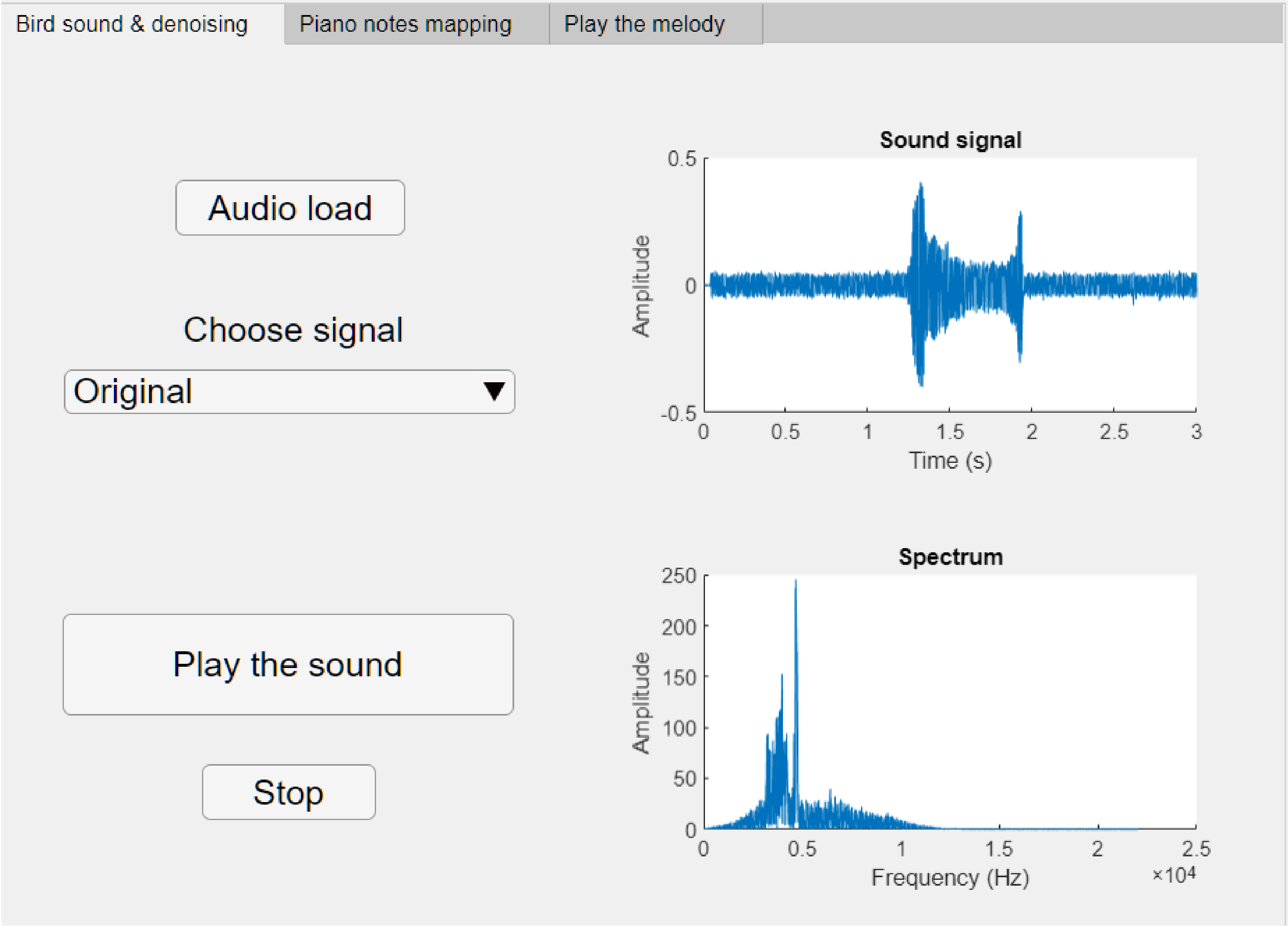
Stage 2 — Melody Integration & demo

Melody

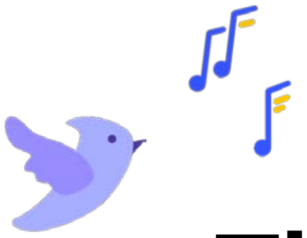
- Connect each single note sound together

Demo

- MATLAB APP Designer
- Let’s play the sound together!



Q & A



Thanks For Your Attention!