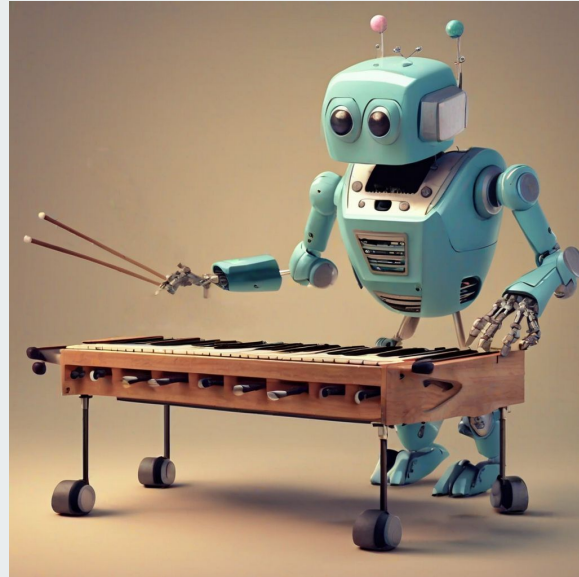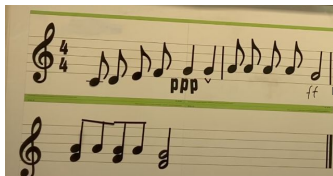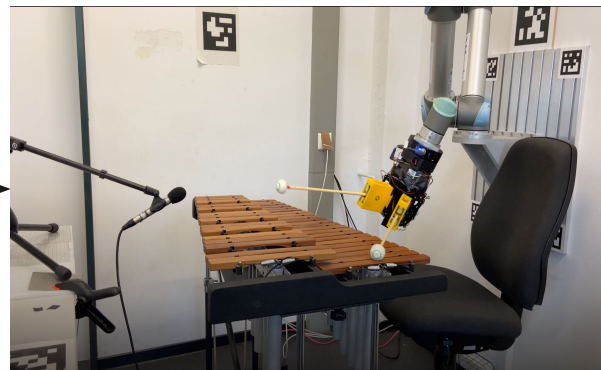# Marimbabot

TAMS Master Project 2022/2023

Task Overview

Sheet Music Input

Robot Behavior

Action

Performance Evaluation

Audio Feedback
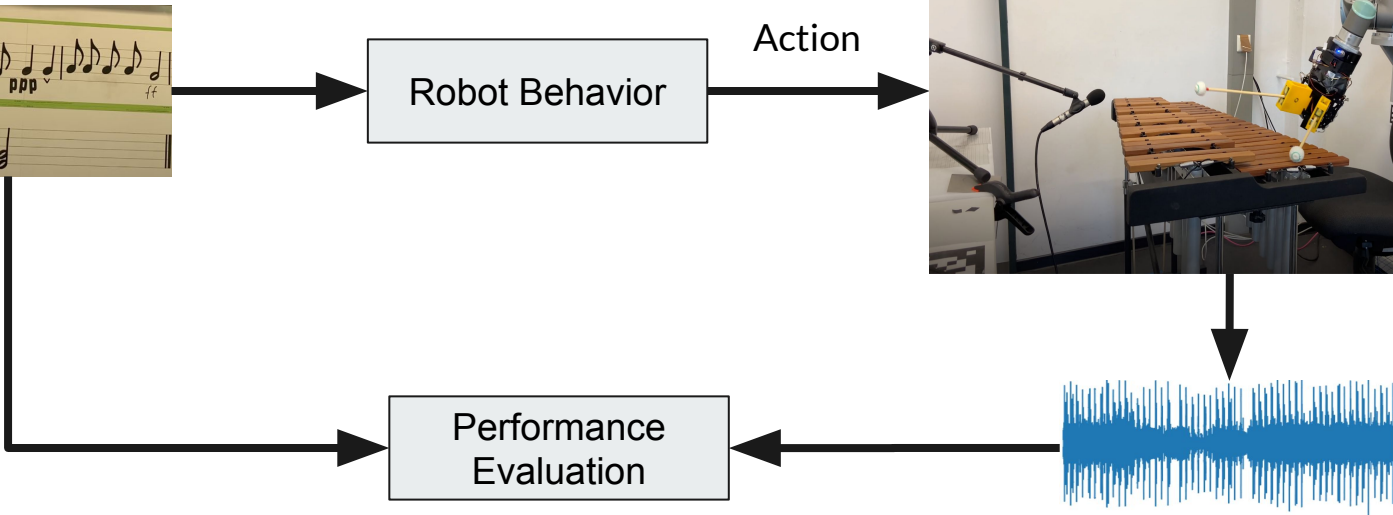
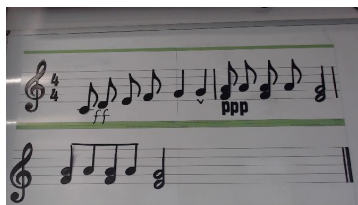# Live Demo

Vision Dataset

Lilypond autogeneration

Hand-written data generation

Manual data collection

Vision Model

Speech

"$100 to 1st place."

normalizer

"one hundred dollars to first place."

phonemizer

wˈʌn hˈʌndɹɪd dˈɑləɹz tə fˈɚst plˈeɪs

phonemes2ids

[1, 0, ..., 0, 2]

voice model

speech synthesis

Audio raw data

Keyword Spotting

Voice Activity Detection

Audio Speech Recognition

Command Extraction
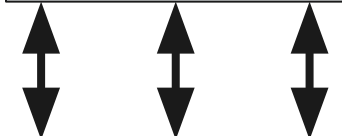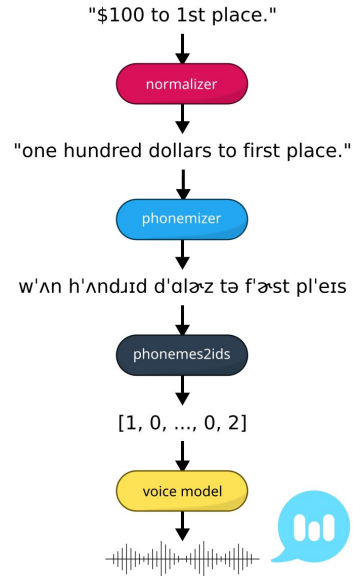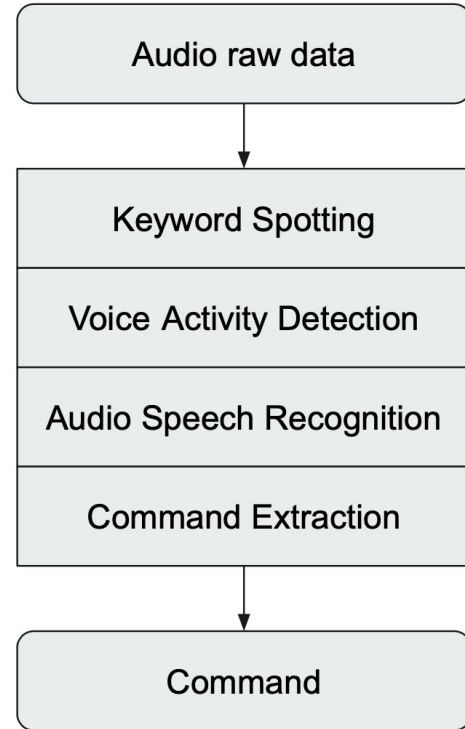
Command

audio command recognition

Behavior
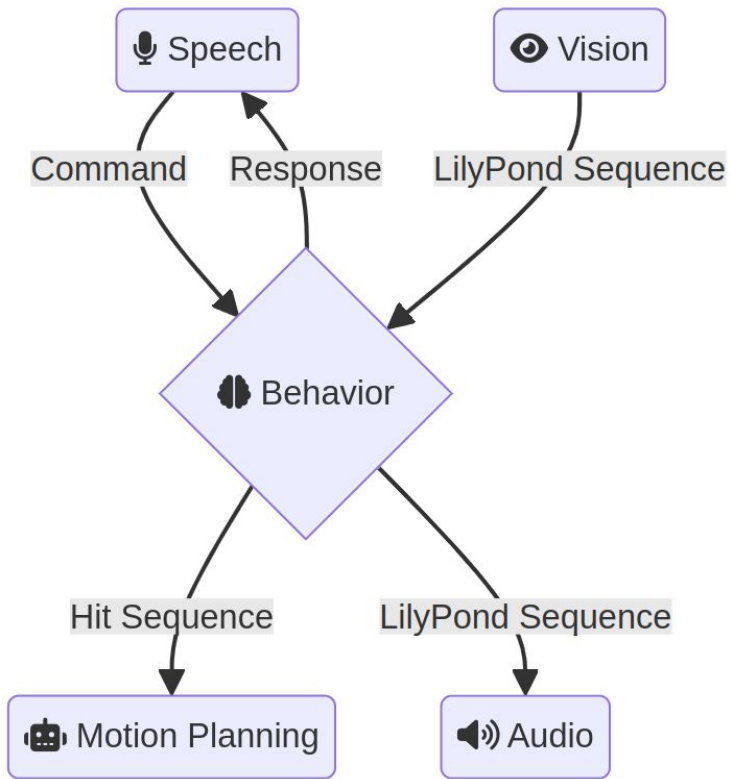
Motion: Three Phase motion

Move to the top of target

Strike motion
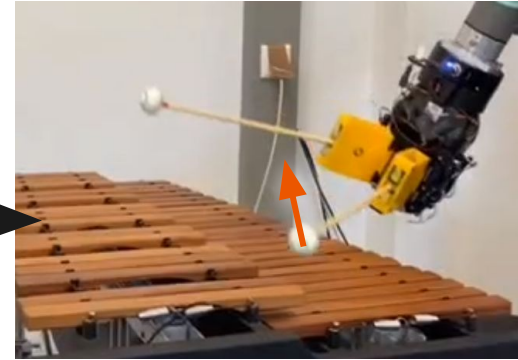
Retreat motion

Simulation

MarimbaBot Simulation Demo
University of Hamburg, TAMS
2023

Hardware

Music Feedback

pitches

2 sec   time frame   0 sec
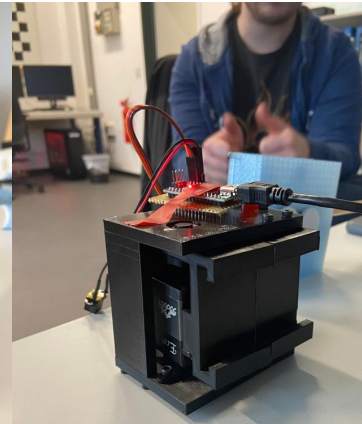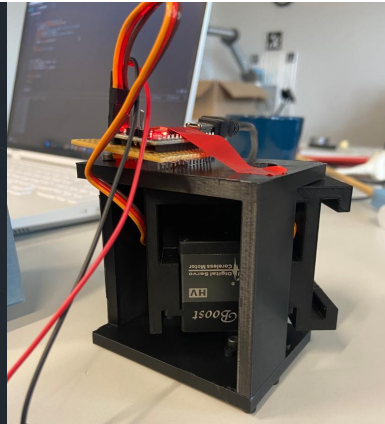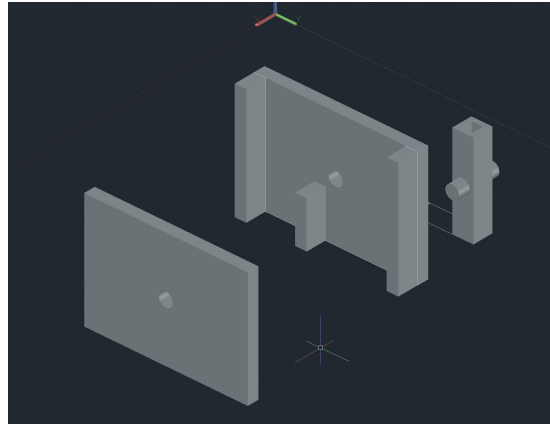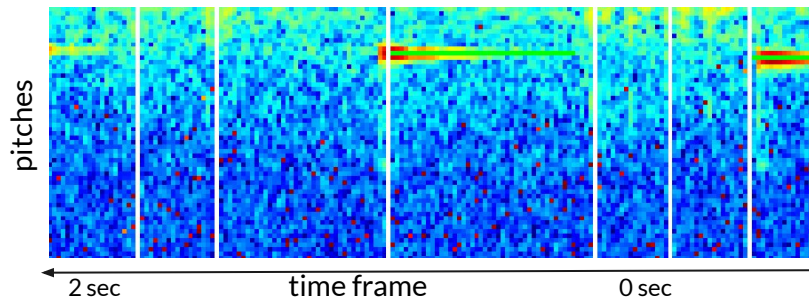
white lines denote candidates
green lines denote the detected notes

Audio raw data

↓

Constant-Q transform

Music Onset Detection

Monophonic Pitch
Classification

↓

Music Note Sequence
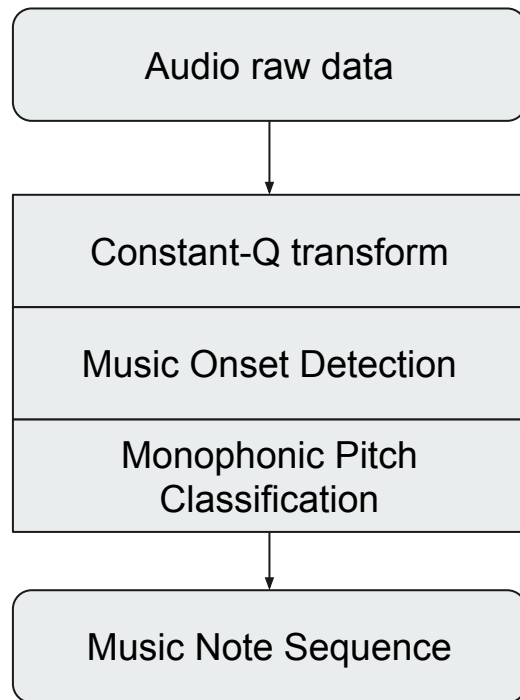
System overview
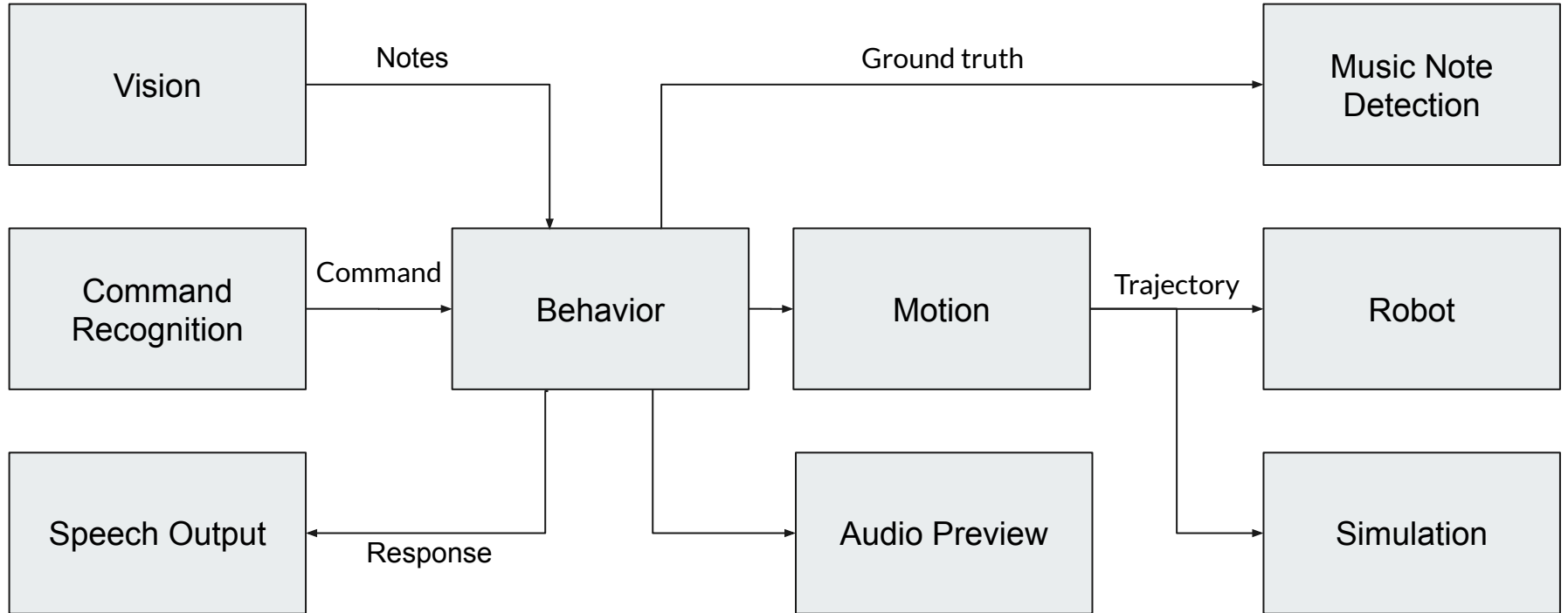
Thank you for your attention !
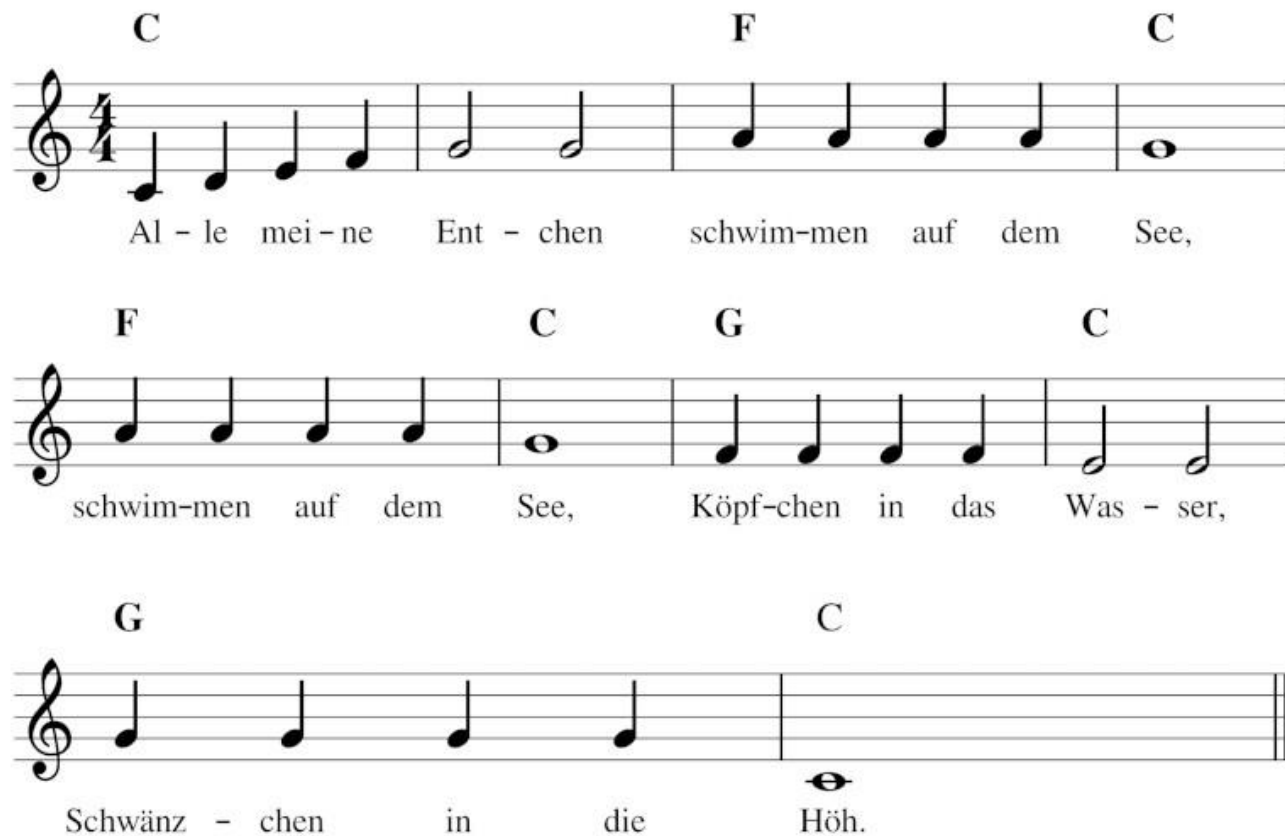
Appendix

Notes

All my Ducklings

Al – le   mei – ne   Ent – chen   schwim-men   auf   dem   See,

schwim-men   auf   dem   See,   Köpf-chen   in   das   Was – ser,

Schwänz – chen   in   die   Höh.

Notes

Frere Jacquès
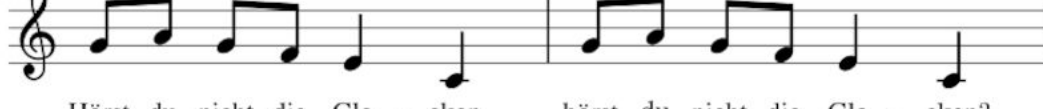
Command Examples

1. Firstly, say "**Hi, marimbabot**" to activate the speech recognition.
2. After you heard the beep, give the command as following examples:

Robot's Action	Command Examples
read the music sheet from white board	"read the music sheet"
start to strike the keys of marimba	"play the music sheet"
increase the playing speed	"play faster (by 20 bpm)"
decrease the playing volume	"play softer (by 2 steps)"
setup the speed	"play in 60 bpm"
repeatedly play the same song	"play in the loop"

Speech

```
┌─────────────────────────────┐
│       Audio raw data        │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│       Keyword Spotting      │
├─────────────────────────────┤
│    Voice Activity Detection │
├─────────────────────────────┤
│  Audio Speech Recognition   │
├─────────────────────────────┤
│     Command Extraction      │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│          Command            │
└─────────────────────────────┘
```

ASR: Whisper

Behavior

Is the Behavior field value 'read'?
- Yes → Receive LilyPond String from the vision module. Success?
  - Yes → Generate corresponding Hit Sequence. Success?
    - Yes → Save LilyPond Sequence as active LilyPond sequence, generate Response: 'Notes recognized.'
    - No → Generate Response: 'No notes recognized.'
  - No → Generate Response: 'No notes recognized.'
- No → Is the Behavior field value 'stop'?
  - Yes → Cancel all Goals.
  - No → Is there an active LilyPond Sequence?
    - Yes → Is there an Action field value?
      - Yes → Modify the LilyPond Sequence accordingly using the default value or Parameter value if given. Success?
        - No → Generate response: 'Unable to apply changes'
        - Yes → Is the Behavior field value 'play'?
      - No → Is the Behavior field value 'play'?
        - Yes → Is the Planning action ready?
          - Yes → Send the active Hit sequence to the Planning action as a Goal. Success?
            - Yes → Is the Action field value 'loop'?
              - Yes → Keep sending the active Hit Sequence as a Goal until stopped.
              - No → Done
            - No → Send response: 'The sequence could not be played.'
          - No → Generate response: 'Robot is already playing.'
        - No → Is the Behavior field value 'preview'?
          - Yes → Is the LilyPond Audio action ready?
            - Yes → Send the active LilyPond Sequence to the LilyPond Audio action as a Goal. Success?
              - Yes → Is the Action field value 'loop'?
                - Yes → Keep sending the active LilyPond Sequence as a Goal unil stopped.
                - No → Done
              - No → Send response: 'The audio preview could not be played.'
            - No → Generate response: 'Preview is already playing.'
          - No → Send response: 'Command not recognized'
    - No → Generate Response: 'No notes to play.'

Electronic Design

Firmware Protocol

Command	Response	Command Description	Response Description
s <value>	ok, err_input_num, err_input_range	Commands the servo to move to the value	Errors for invalid number formats or numbers out of range of the safety limits or ok for acknowledgement
p	p <value>	Requests the current target value	Returns the current target value
l	l <min_val> <max_val> <val_resolution>	Requests the limits of the device	Return the max and min value bounds and the resolution of the value parameter (0-res)
<invalid_command>	err_cmd	Any command other than the listed ones	Error code for and invalid / unknown command