

# FuFeeling

## 语音情绪感知伴侣宠物

第四组

# 目录

01

产品背景

02

产品介绍

03

产品原理

04

未来展望







PART 01

产品背景

Product Background

# 陪伴机器人的兴起



## 孤独成为某种生活状态



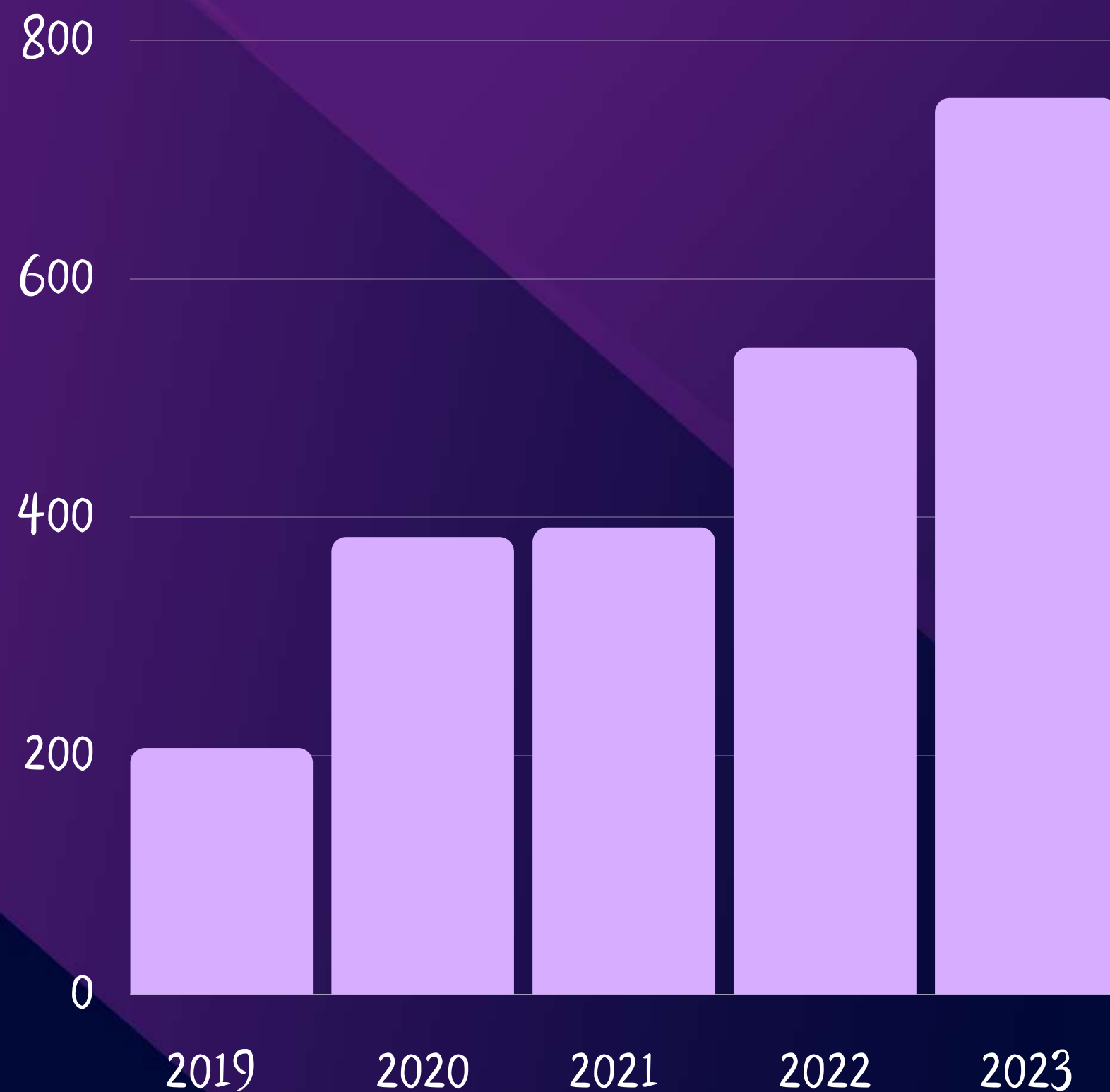
2022年我国独居人口数量破亿，独居率超30%  
2021年底，全国60岁及以上老年人口达2.67亿，  
“空巢老人”“空巢青年”现象急剧升高

## “群体性孤独”下的情绪问题



情绪障碍中抑郁症接受治疗的比率仅为9.5%  
大多数选择以自我调节的方式应对心理困扰

# 陪伴机器人市场规模



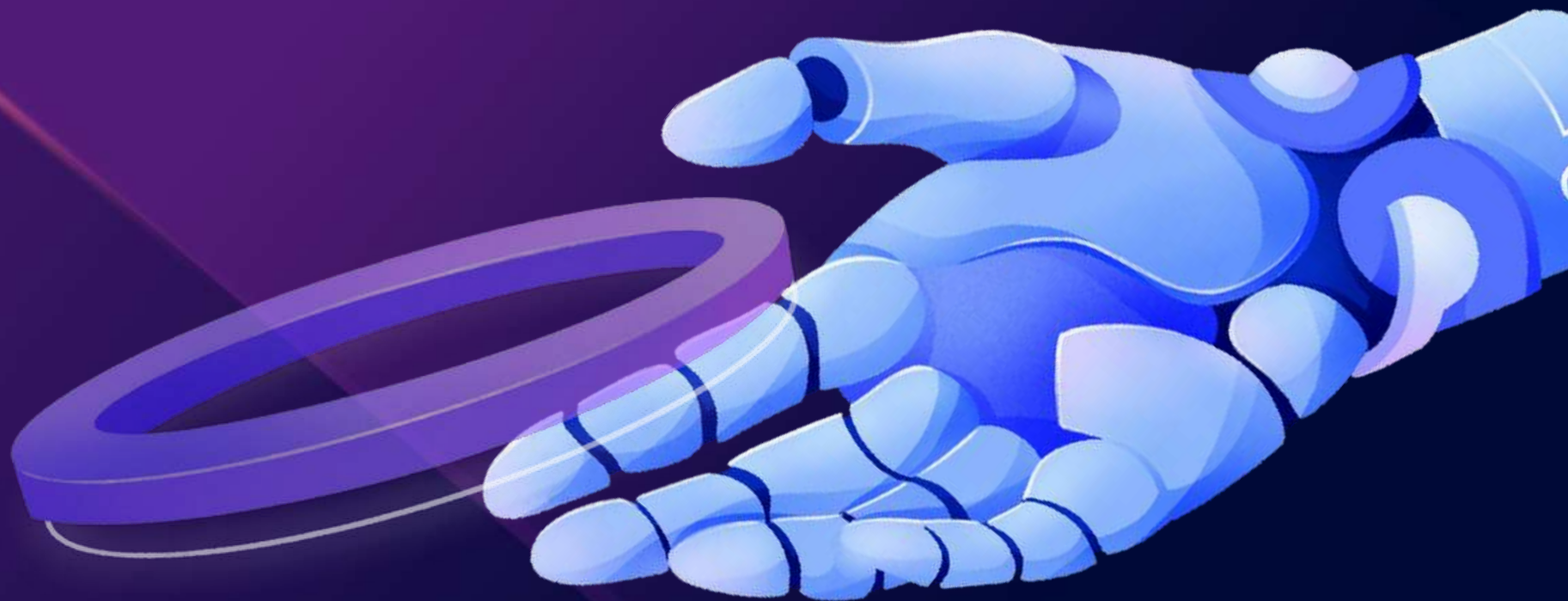
服务型机器人市场占比到达  
781亿元左右

其中陪伴机器人的关注度占比  
超过40%以上

日本Groove X公司、以色列  
Intuition Robotics公司、英国  
Consequential Robotics公  
司等已经投入开发



# 陪伴机器人： 数字治疗实现心理治疗



## 天生的私密感

人工智能的好处在于它不会移情，让人有信赖感，可以直接打开话题，因为对方知道这就是一个机器人，让人感觉到天生的私密感/好感，吐露心声

## 主动提出帮助

目前大多数的情感障碍或焦虑抑郁人群不会主动寻求心理咨询帮助，而作为机器人则可以观察到细微的情绪变化从而提出帮助，更易取得安抚的效果

## “心灵维C”式陪伴

它24x7随叫随到，不用预约、不用出门、不用担心深夜吵醒朋友、也不用担心被主观地价值判断，细水长流式的陪伴和安抚具有其他心理咨询没有的优势





# 技术难点——情感计算

01

## 识别

语音、语言、表情和姿态等从不同的维度精确捕捉人类的情感表达

02


## 决策

纯理性的决策做出的反应很难得到共鸣，用情感来回馈情感加入决策模型

03

## 表达

通过语音、肢体和表情，机器人对人做出不那么具有情感温度的回馈表达





# PART 02

## 产品介绍

Product Introduction



# 产品展示

我们基于**语音记录**，提取语音中的各种声音特征来实时判断陪伴人的情绪，并将情绪信息接入人工智能**大模型**来生成文字反馈，并合成拟人化语音输出。



# 演示视频

主演： 友情出演： 拍摄： 后期：



PART 03

# 技术原理

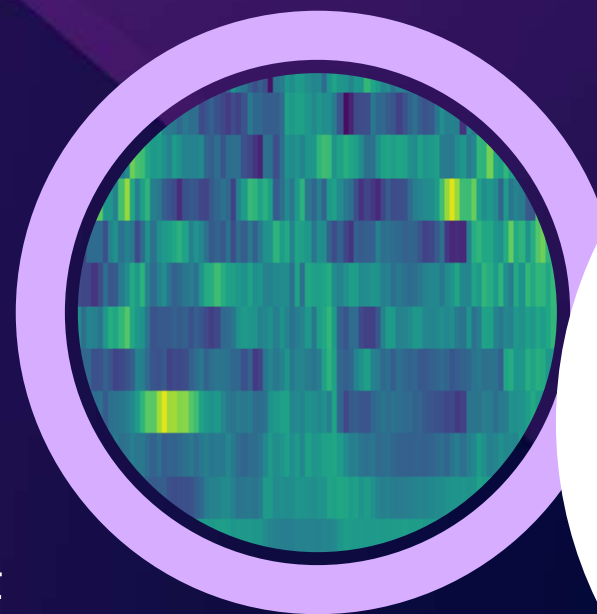
Technology and Principle



# 总体内容框架

## MFCC算法

通过提取情绪音频的MFCC特征来分类音频，总共识别“高兴”，“伤心”，“生气”以及环境音四种声音。



## 神经网络

获得MFCC特征向量后，我们利用卷积神经网络进行特性参数训练以及推测

## 大模型

我们将整个产品接入了科大讯飞星火认知大模型，用于生成情绪反馈以及语音合成。



# 数据采集

我们的数据集采用两个来源：

- 1、自己采集160条不同情绪的语音（文字无意义）
- 2、通过数据检索寻找到类似工作，向其作者Pan Liu, PhD（NORTH DAKOTA STATE UNIVERSITY）获取共450条数据集。



Dear professor,

Hello!

I'm an undergraduate in Zhejiang University, China.

I've just read your great article "Recognizing vocal emotions in Mandarin Chinese: A validated database of Chinese vocal emotional stimuli", and I'm doing the related research. Could you please share your dataset with me?

Thank you very much!

Yours sincerely,

Li Chenrui



**Pan Liu** <pliu261@gmail.com>

to me ▾

Mon, Oct 16, 9:53 PM ☆ ↶ ⋮

Here is the database: <https://uwoca-my.harvard.edu/>



Let me know if you have any problems accessing it.

...

**Pan Liu, PhD** (she/her/hers)

Assistant Professor

Department of Psychology

**NORTH DAKOTA STATE UNIVERSITY**

134D8 Minard Hall, 1210 Albrecht Blvd, Fargo ND 58102

701-231-5676

[pan.liu@ndsu.edu](mailto:pan.liu@ndsu.edu)

[panliu.org](http://panliu.org)

# MFCC算法



## 基本原理

MFCC特征是一种基于内耳频率分析的人类声音感知模型，MFCC 集提供了具有感知意义的，平滑的语音频谱随时间的估计。人类内耳结构工作原理：机械震动在耳蜗的入口产生驻波，引起基底膜以与输入声波频率相称的频率协调在此频率上的最大幅度震动。

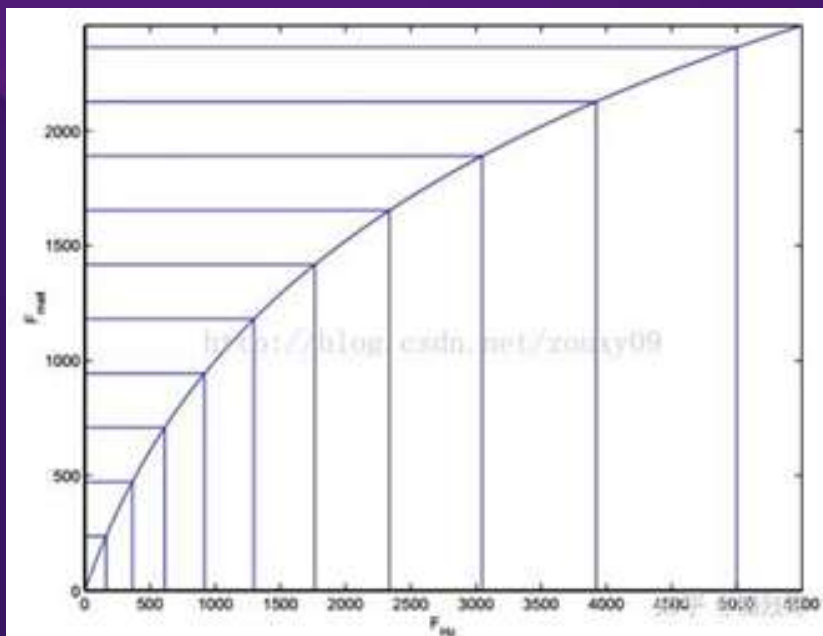


## 实现过程

1. 将音频信号分帧加窗
2. 利用傅立叶变换计算频谱
3. 将频谱按照 Bark 的划分方式划分为 N 个单元，对这 N 个单元使用滤波器组分别计算每个滤波器对应的 Bark 单元中的能量
4. 将 3 中获取的每个能量转换为对应的对数能量
5. 对 4 中的结果作离散余弦变换（DCT）
6. 保留变换后的 12~20 个结果，该结果即为求解的 MFCC 特征向量
7. 特征归一化



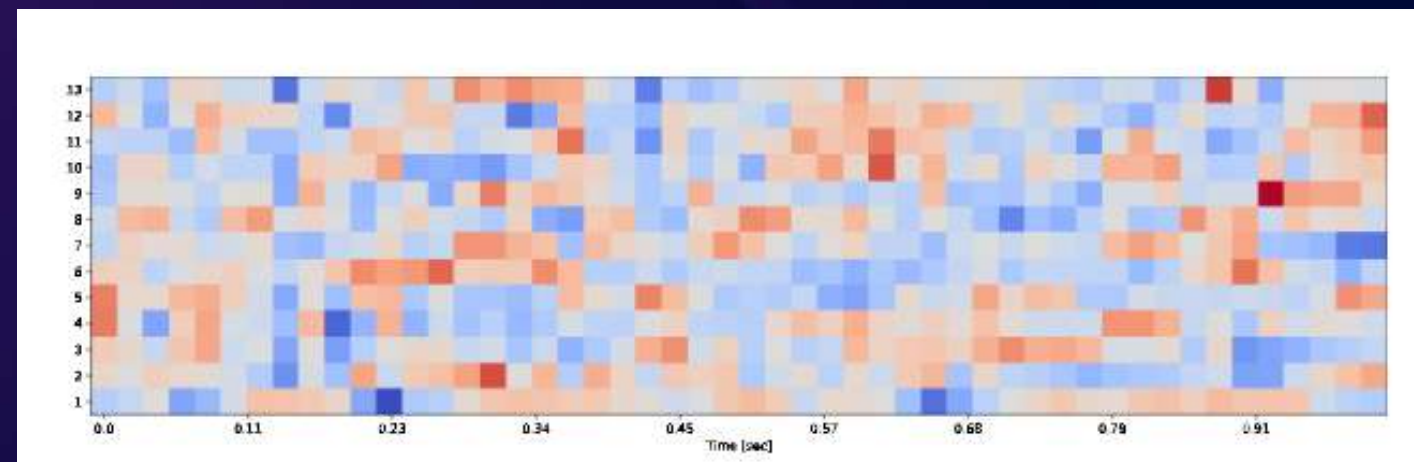
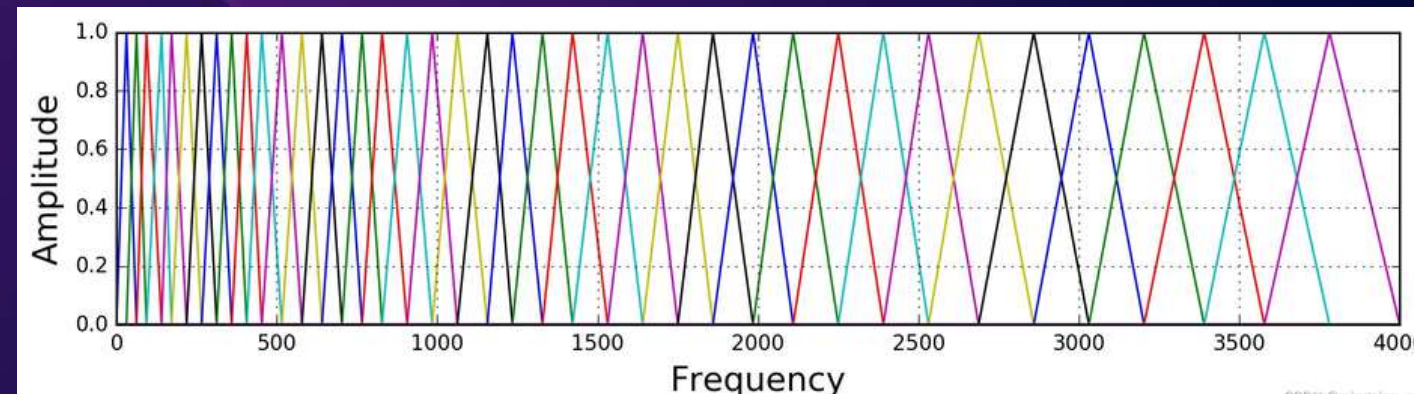
# 梅尔滤波



$$M(f) = 1125 * \ln(1 + f/700)$$

## 梅尔尺度(Mel Scale)

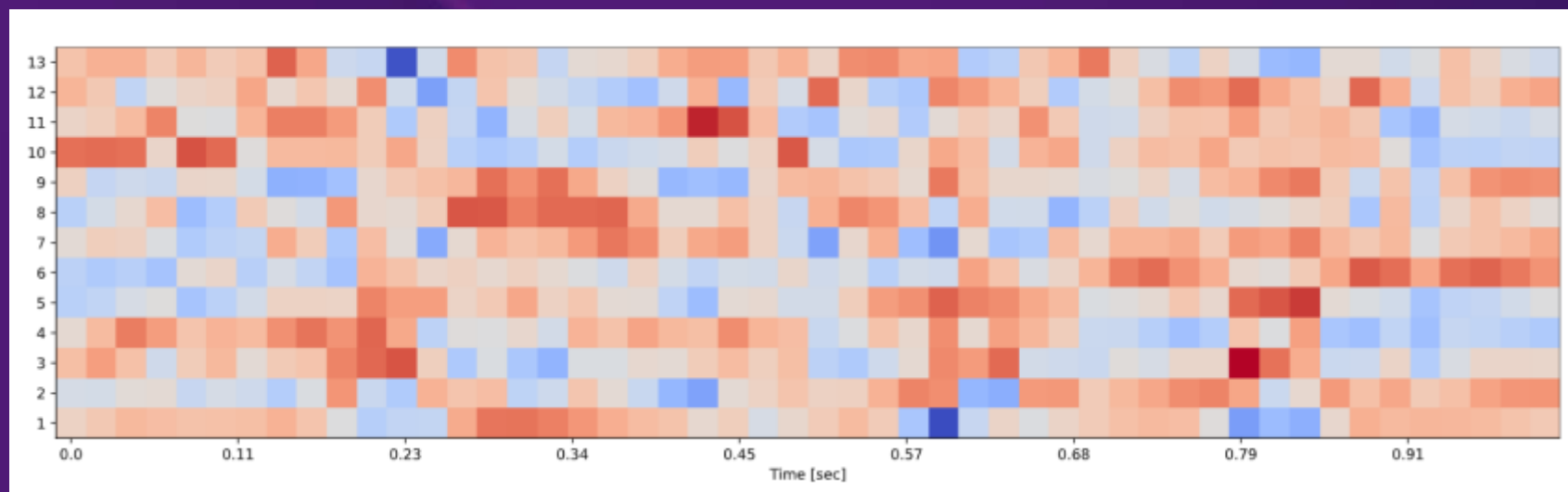
美尔尺度是建立从人类的听觉感知的频率——Pitch到声音实际频率直接的映射。人耳对于低频声音的分辨率要高于高频的声音。通过把频率转换成美尔尺度，我们的特征能够更好的匹配人类的听觉感知效果。



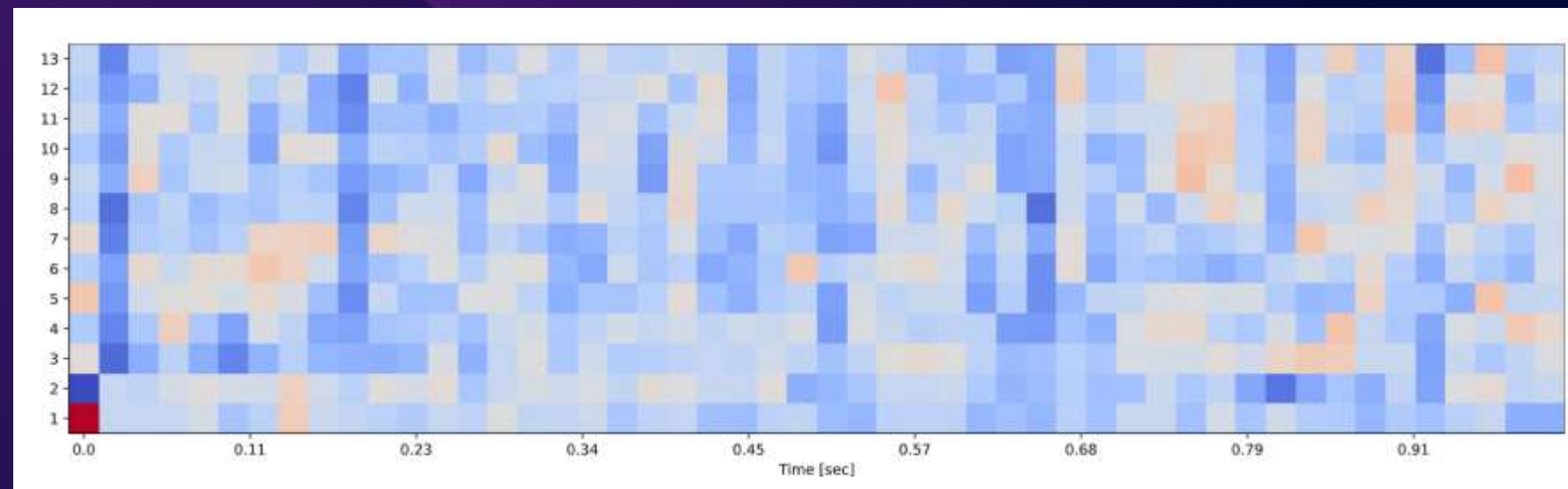
## 梅尔滤波器组

采用了16个三角滤波器，经过对数化、DCT后得到12个倒谱系数，作为每条音频的特征。

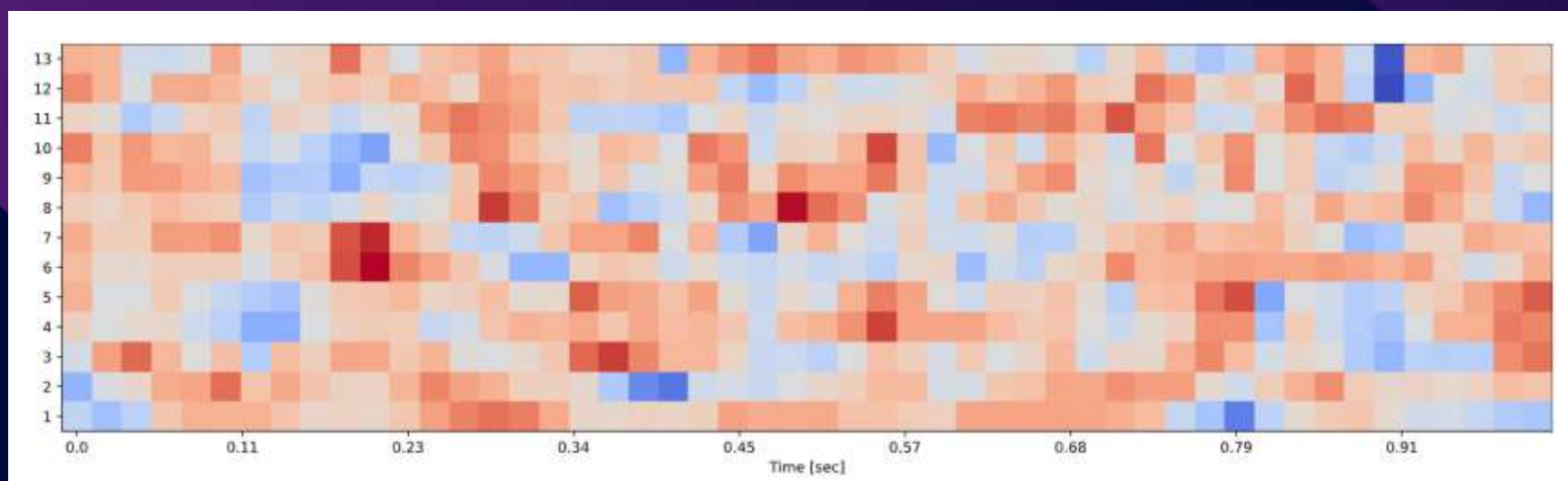
# 计算出的MFCC系数



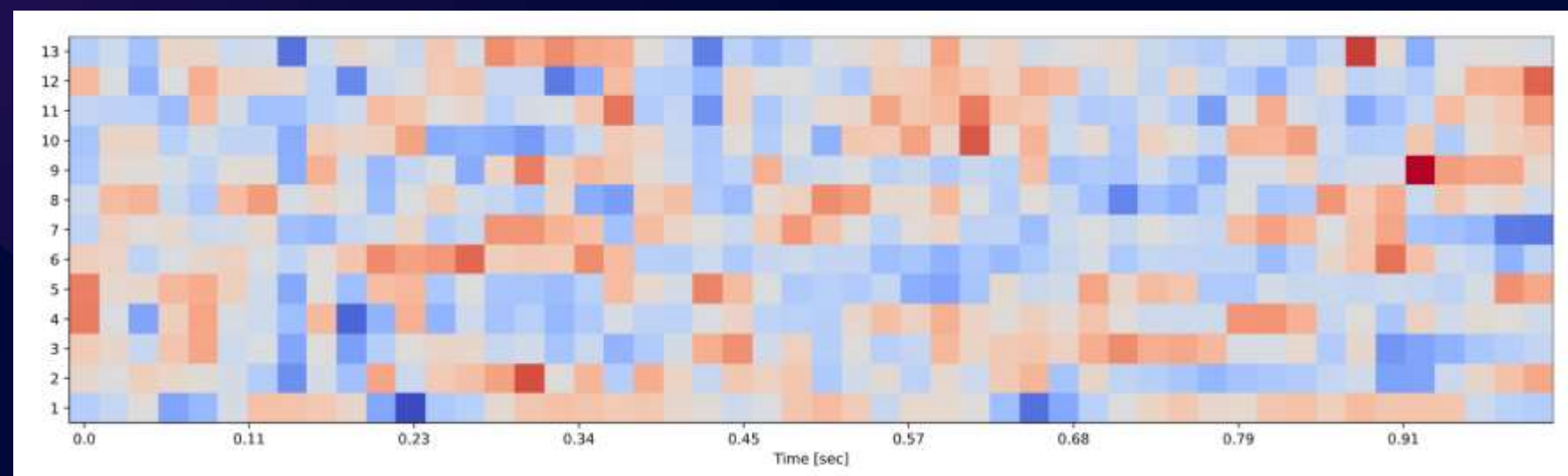
HAPPY



ENVIRONMENT



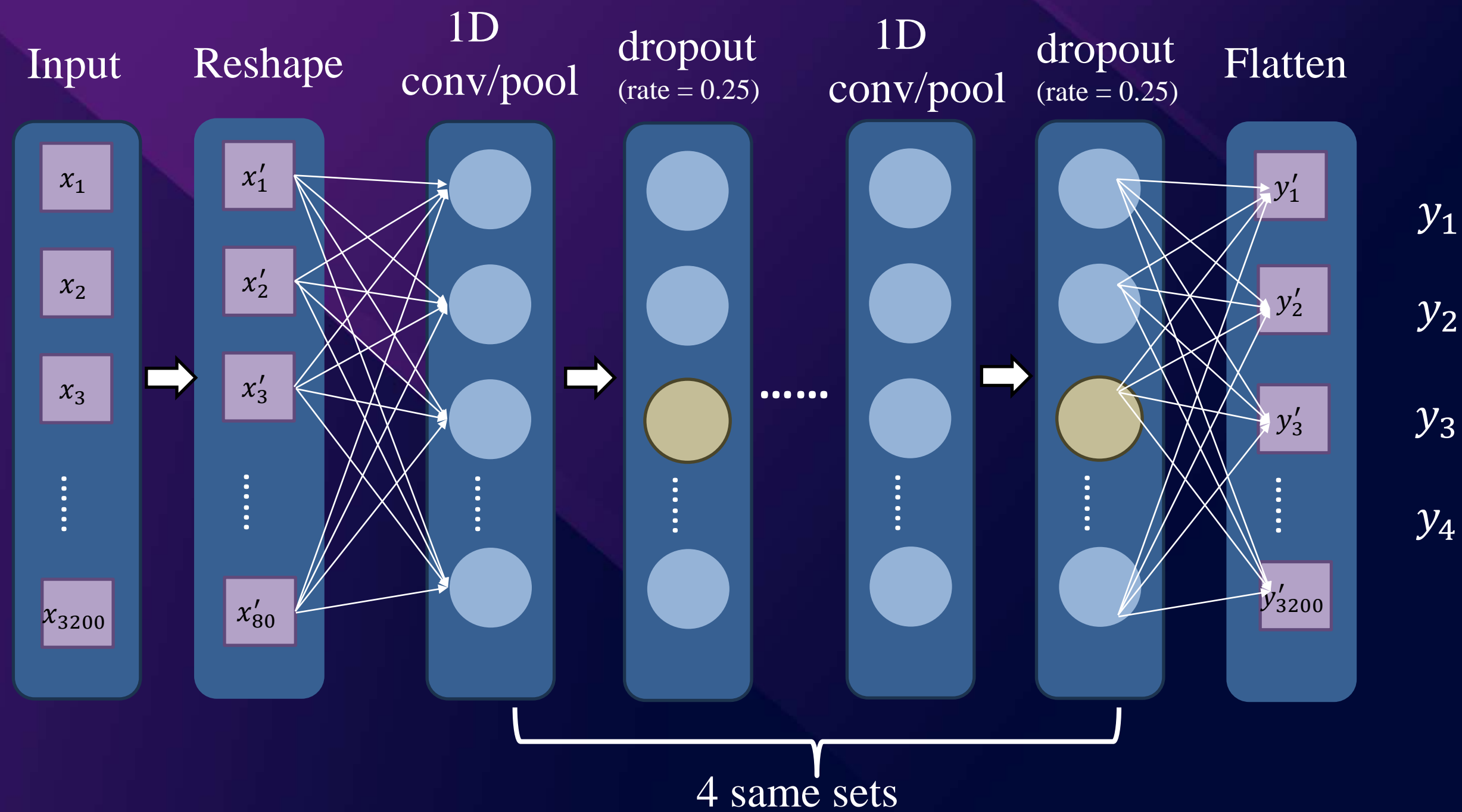
SAD



ANGRY



# 神经网络





# Python串口通信

利用Serial库进行Arduino与Python在串口的通信，以对Arduino上的情绪检测结果返回python传入大模型

```
void setup() {  
    // put your setup code here, to run once:  
    Serial.begin(115200);  
}  
  
void loop() {  
    // put your main code here, to run repeatedly:  
    Serial.print(1);  
    delay(110000);  
}
```

```
import serial  
  
#串口通信设置  
serialPort1 = "COM10"  
baudRate = 115200  
ser1 = serial.Serial(serialPort1, baudRate, timeout=0.5)  
#print("参数设置: 串口=%s , 波特率=%d" % (serialPort, baudRate))  
  
while(1):  
    #Input = input("\n" + "我:")  
    data_dec1 = ser1.read_all()  
  
    if data_dec1:  
        rec_str1 = data_dec1.decode('utf-8')  
        print(data_dec1)
```

# 星火大模型API调用



## 文本反馈

Arduino板通过串口向Python发送情绪判断结果，根据结果程序调用星火大模型3.0专用API接口上传结果并获得文字信息输出



## 语音合成

将大模型反馈回的语音进行进行储存，调用另外的语言合成模型API专用接口，返回MP3文件并即时播放

```
#以下密钥信息从控制台获取
appid = "90...c42" #填写控制台中获取的 APPID 信息
api_secret = "MTBiMW...TNhYjAxMDQyOGI1" #填写控制台中获取的 APISecret 信息
api_key = "2828c...58f24...33c7c4d" #填写控制台中获取的 APIKey 信息

#用于配置大模型版本，默认"general/generalv2"
domain = "general" # v1.5版本
# domain = "generalv2" # v2.0版本
#云端环境的服务地址
Spark_url = "ws://spark-api.xf-yun.com/v1.1/chat" # v1.5环境的地址
# Spark_url = "ws://spark-api.xf-yun.com/v2.1/chat" # v2.0环境的地址
```

```
class Ws_Param(object):
    # 初始化
    def __init__(self, APPID, APIKey, APISecret, Text):
        self.APPID = APPID
        self.APIKey = APIKey
        self.APISecret = APISecret
        self.Text = Text

    # 公共参数(common)
    self.CommonArgs = {"app_id": self.APPID}
    # 业务参数(business)，更多个性化参数可在官网查看
    self.BusinessArgs = {"aue": "lame", "auf": "audio/L16;rate=16000", "vcn": "x4_doudou", "tte": "utf8"}
    self.Data = {"status": 2, "text": str(base64.b64encode(self.Text.encode('utf-8')), "UTF8")}
    #使用小语种须使用以下方式，此处的unicode指的是 utf16小端的编码方式，即"UTF-16LE"
    #self.Data = {"status": 2, "text": str(base64.b64encode(self.Text.encode('utf-16')), "UTF8")}
```

# 总体实现路线图



预加重、分帧、加窗

FFT

取绝对值或平方值

Mel滤波

取对数

DCT

动态特征



1D卷积神经网络

大模型Web API调用

语音合成 API调用







PART 04

未来展望

Future Prediction

# 未来方向

01

## 多模融合识别

目前的识别只基于语音，在准确度、应用范围上有限制，未来可以开发更多如光电等检测模式融合识别

02

## 更能共情的反馈

目前的反馈较为生硬，没有做到对现实情况的针对性反馈，未来将细化反馈条件，使得反馈更加智能

03

## 持续性反馈

目前只能做到一次性的反馈，当反馈开始后无法连贯性的开启对话。未来接入语义识别功能，能加陪伴性的对话。



# 小组分工

## NAME

## CONTRIBUTION

A

python串口通信、API调用、视频拍摄、  
PPT制作、展示

B

模型数据集获取、训练和调试、视频演  
出、报告编写、展示

C

Arduino代码编写、视频后期

D

模型数据集获取、训练和调试、视频演  
出、报告编写、Arduino代码编写





Thanks for  
Watching

谢谢  
观看

第四组：



如果没猜错的话现在是上课时间吧，同学