开源声码器WORLD在语音合成中的应用

喜马拉雅FM音视频高级工程师

马力

begeekmyfriend@gmail.com

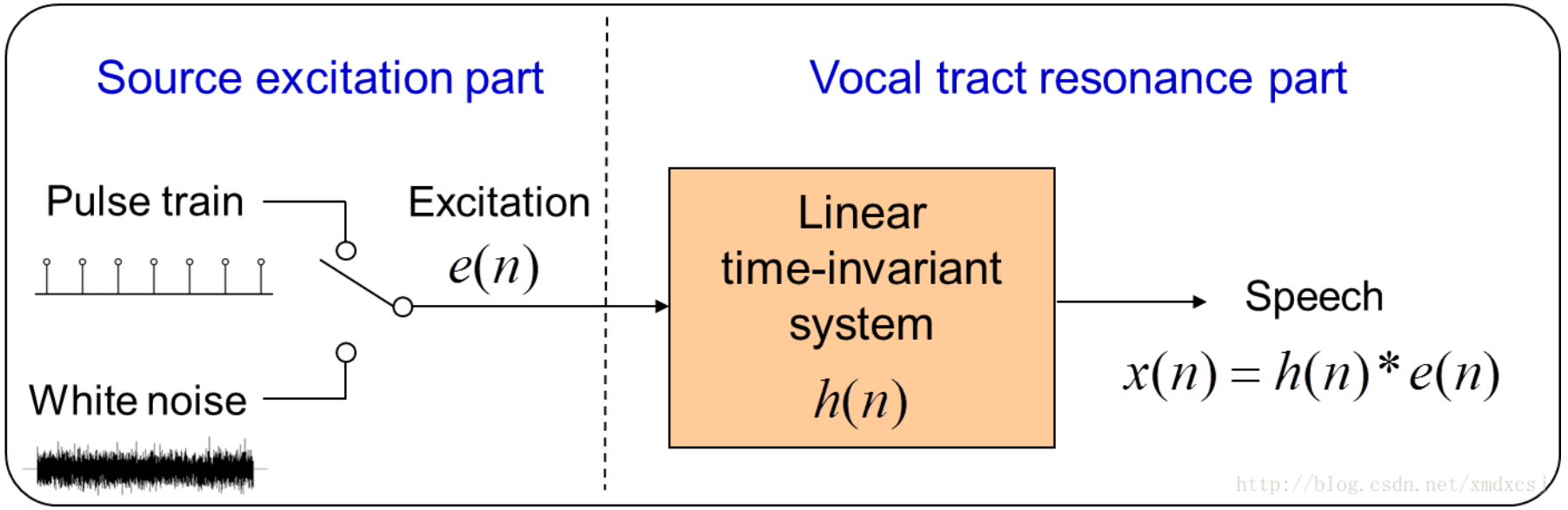
Tacotron + WORLD

- 关于西藏的传说有很多，历来都是朝圣者的天堂。而作为中国西南边陲重地，也都是中国领土不可分割的一部分。二零一五年，央视曾经播出过一部高分纪录片《第三极》。片中，天高地阔的风景，让无数人对西藏“情根深种”。时隔两年，由原班人马打造的姊妹篇《极地》悄然上线！每一帧都是壁纸，每一幕都是人间仙境。自影片播出之来好评如潮，就连一向以严谨出名的豆瓣评分也是很高。早在二零一五年，它的第一季《第三极》就拿到了豆瓣9.2分。而让它一下拿到9.5分的原因，是因为它展示了在那片绝美与贫瘠并存的净土上，普通人的真实生活是什么样子。

What does a vocoder do?

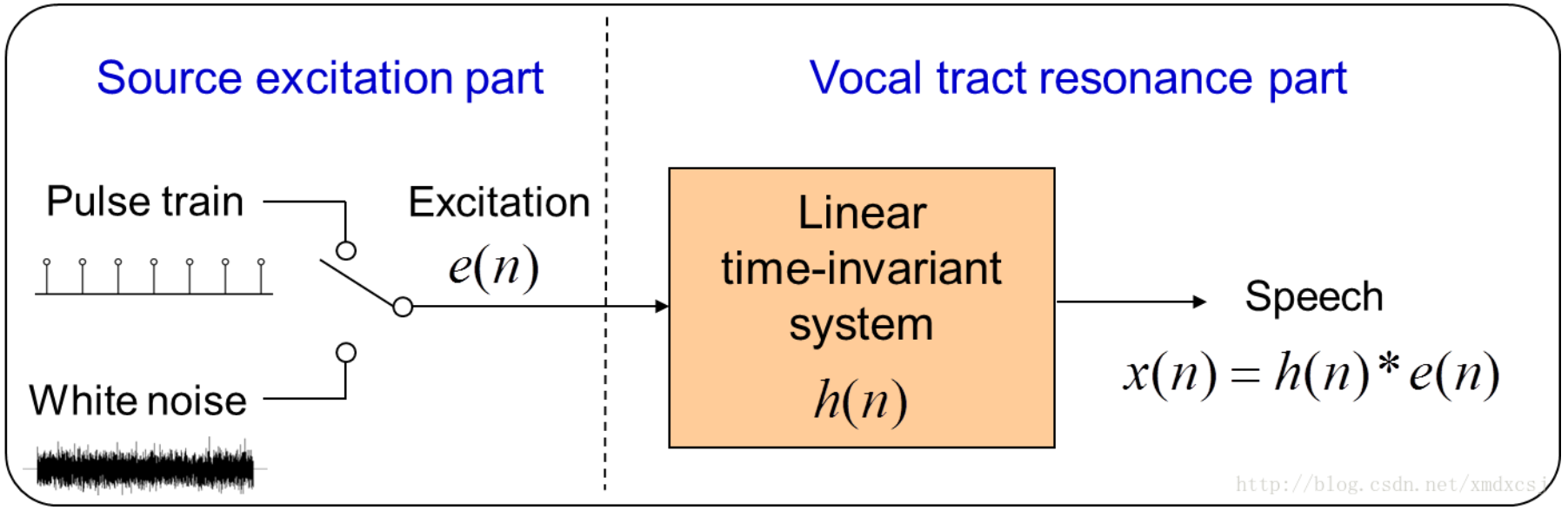
- Acoustic Features
  - Analysis
  - Manipulation
  - Synthesis

Acoustic Modeling



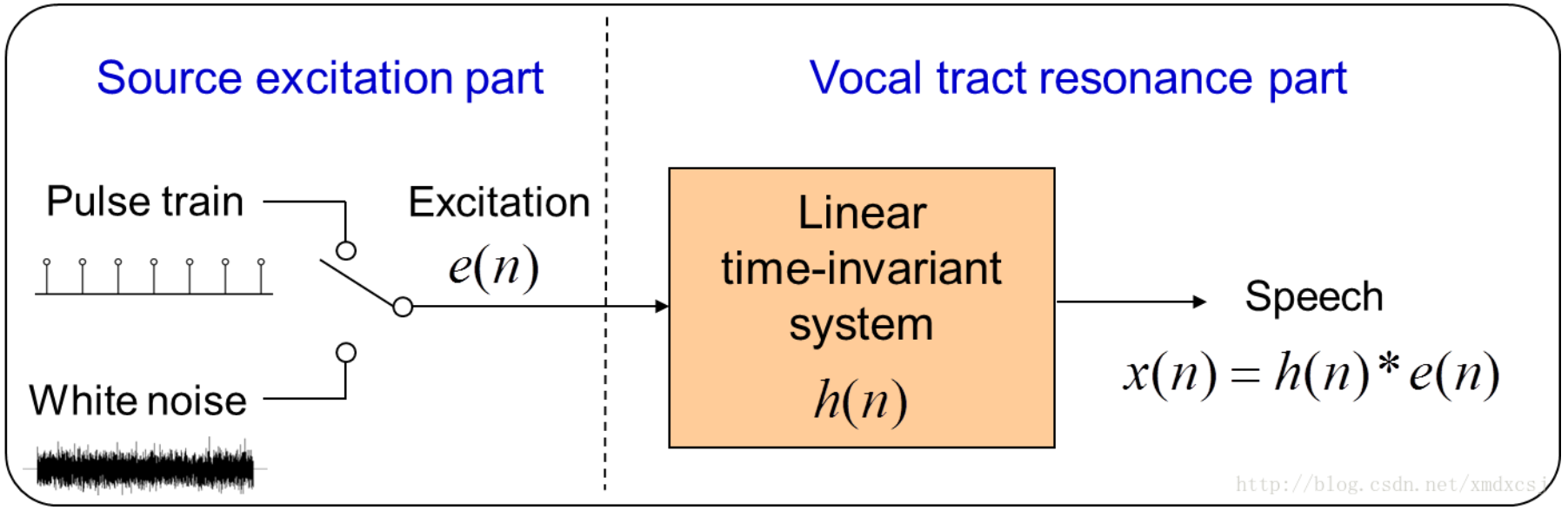
人发声机理的经典**源-滤波器**（source-filter）模型
源激励部分对应于肺部的气流和声带共同作用形成的激励
声道谐振部分对应于声道的调音运动。

Acoustic Modeling



浊音(voiced)：气流通过紧绷的声带，对声带进行冲击而产生振动，使声门处形成准周期性的脉冲串，激励信号简化为**周期性的脉冲激励**。
清音(unvoiced)：声带处于松弛状态，不发生振动，气流通过声门直接进入声道，激励信号简化为**随机白噪声**。

Acoustic Modeling



F0基频对应激励部分周期脉冲序列；
SP频谱包络对应声道谐振部分时不变系统冲激响应；
AP非周期序列对应**混合激励部分**非周期脉冲序列。

Acoustic Features

- F0(fundamental frequency)
  - 一组正弦波组成原始信号，频率最低的正弦波为基频，其它为泛音。
- SP(spectral envelop)
  - 将不同频率的振幅最高点通过平滑的曲线连接起来就是频谱包络。
- AP(aperiodicity)
  - 语音的非周期信号参数
  - 混合激励可以通过aperiodicity来控制浊音段中的周期激励和噪声（非周期）成分的相对比重

WORLD vocoder

- Sound quality
- Processing speed
- Open source

WORLD Analysis

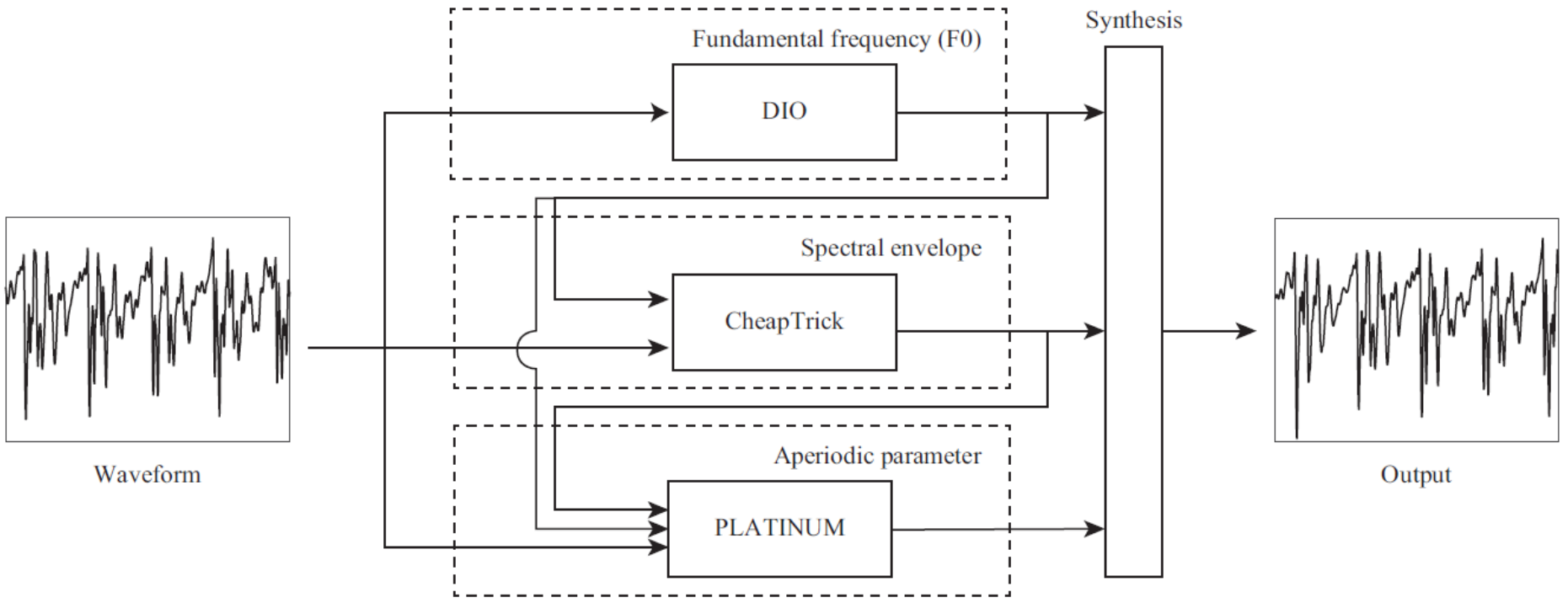


Fig. 1 Overview of the developed system. WORLD consists of three analysis algorithms for determining the F0, spectral envelope, and aperiodic parameters and a synthesis algorithm incorporating these parameters.

Fundamental Frequency

- DIO
  - 低通滤波器对原始信号滤波
  - 取四个周期计算置信度（标准差）
    - zero-crossing intervals
    - peak intervals
    - dip intervals
  - 选取标准差最低的作为基频

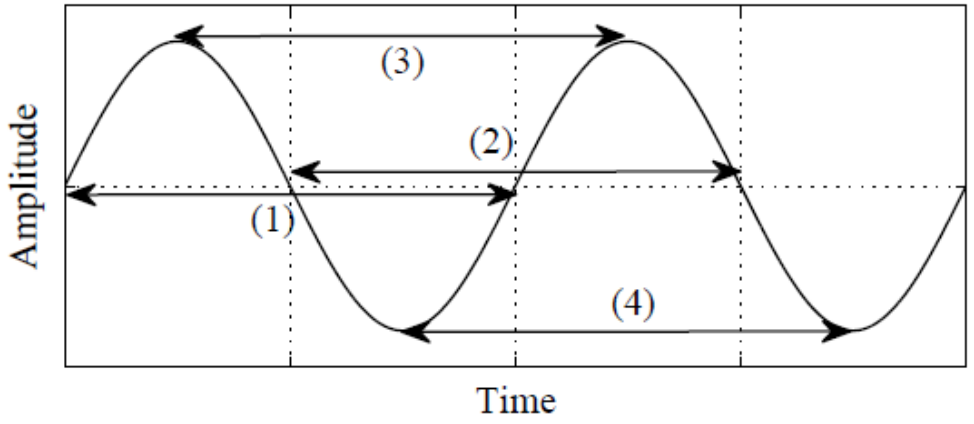
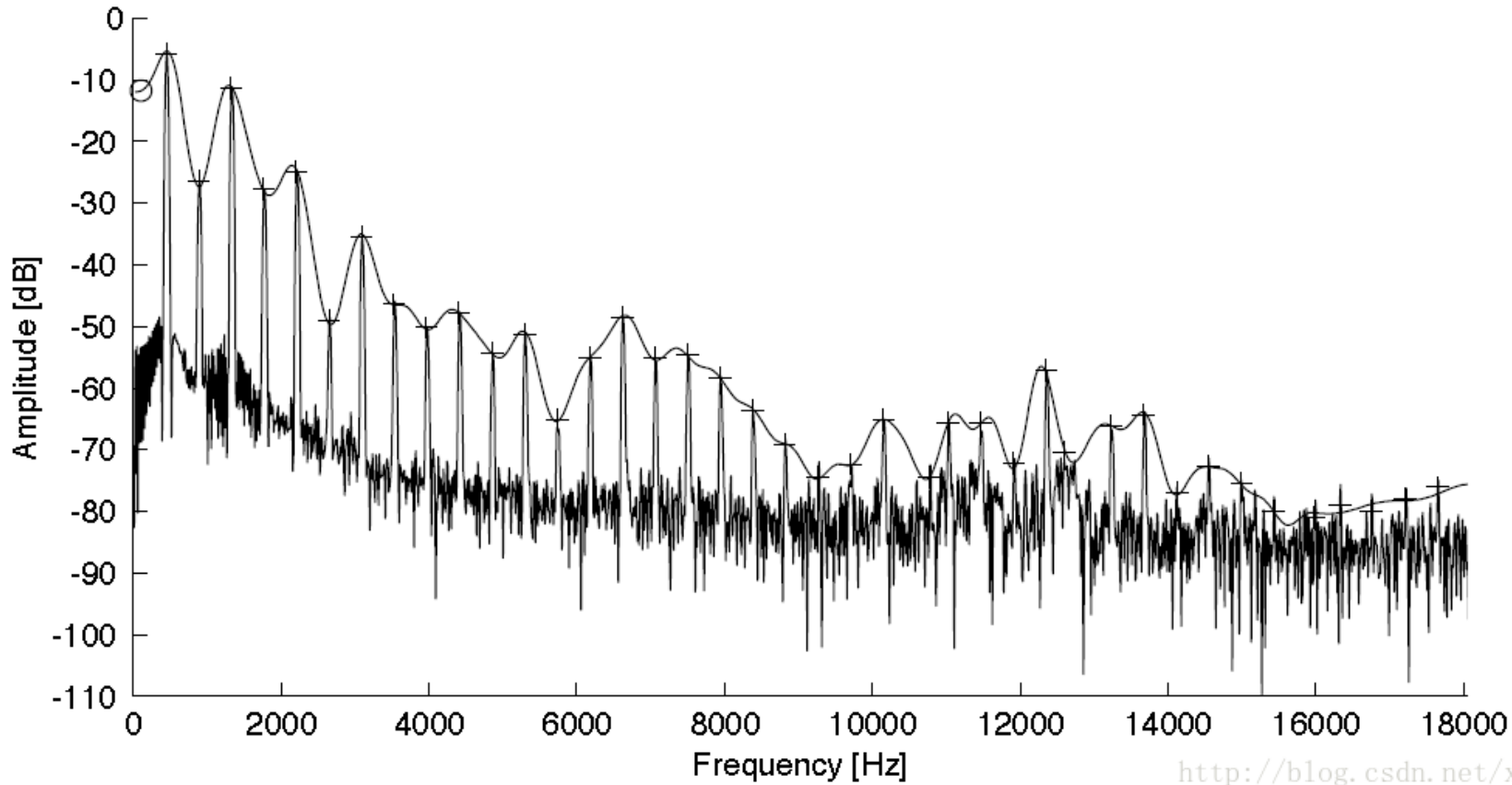


Fig. 2 Four intervals used for calculating an F0 candidate and its reliability. If the filtered signal only consists of the fundamental component, the four intervals indicate the same value.

Spectral Envelop



Spectral Envelop

- LPC
  - 一个语音取样值可以用若干个语音取样过去值的**加权线性组合**来逼近
- Cepstrum
  - 复数倒谱拥有频谱**幅度**跟**相位**的信息
  - 信号 -> FFT -> 取绝对值 -> 取对数 -> 相位展开 -> IFFT -> 倒频谱
- CheapTrick (pitch synchronous analysis)
  - F0 adaptive windowing
  - Smoothing of the power spectrum
  - Liftering in the quefrency domain

Aperiodicity

- PLANTINUM
- D4C
  - 计算群延迟
  - 修正参数
  - 估计band-aperiodicity

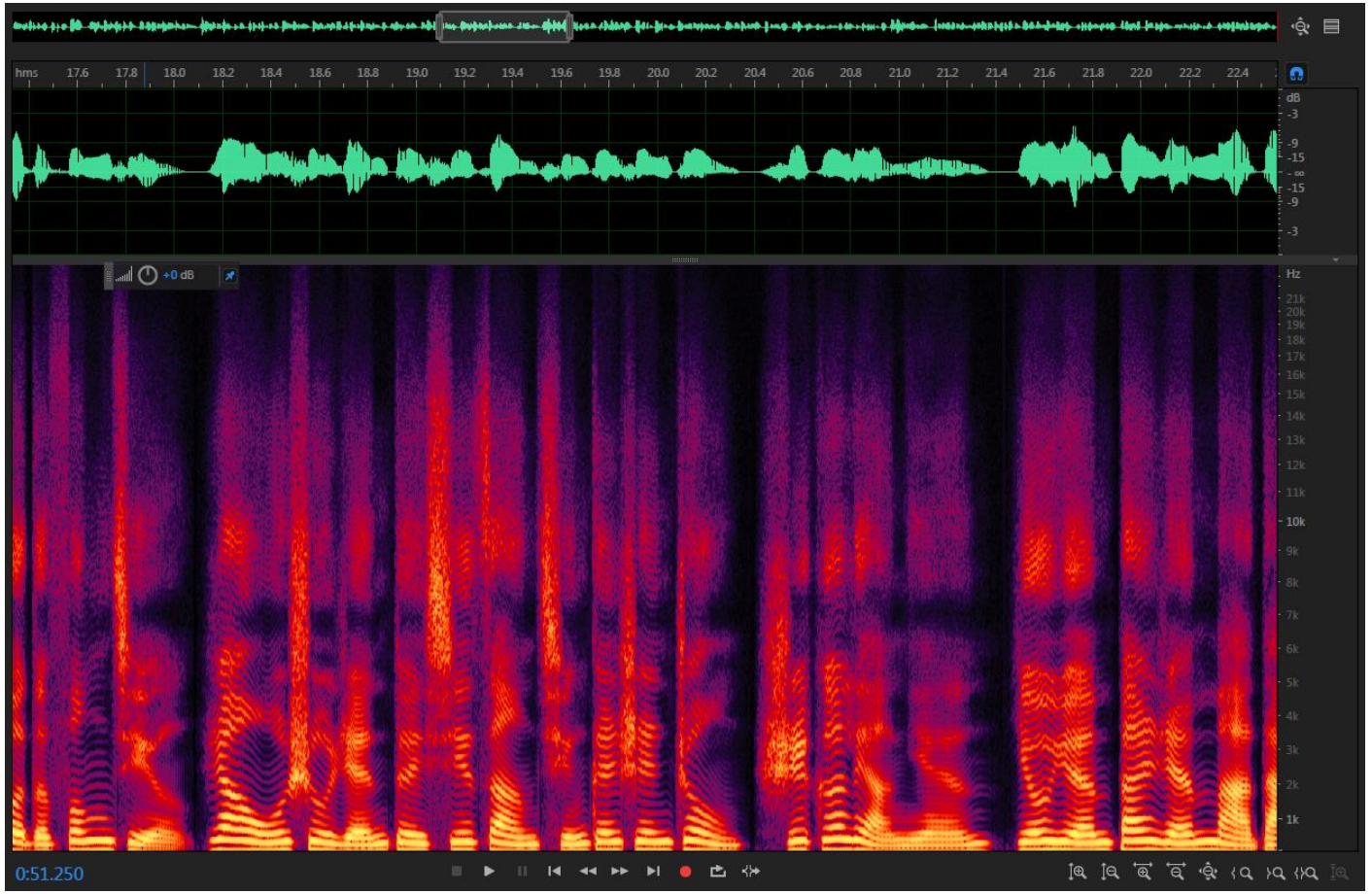
Python Wrapper

```
        cpp_aperiodicity[i] = &aperiodicity0[i, 0]

    Synthesize(&f0[0], f0_length, cpp_spectrogram,
        cpp_aperiodicity, fft_size, frame_period, fs, y_length, &y[0])
    return y

def wav2world(x, fs, fft_size=None, frame_period=default_frame_period, ap_depth=5):
    """Convenience function to do all WORLD analysis steps in a single call.

    In this case only `frame_period` can be configured and other parameters
    are fixed to their defaults. Likewise, F0 estimation is fixed to
    DIO plus StoneMask refinement.

    Parameters
    ----------
    x : ndarray
        Input waveform signal.
    fs : int
        Sample rate of input signal in Hz.
    frame_period : float
        Period between consecutive frames in milliseconds.
        Default: 5.0
    fft_size : int
        Length of Fast Fourier Transform (in number of samples)
        The resulting dimension of `ap` adn `sp` will be `fft_size` // 2 + 1

    Returns
    -------
    f0 : ndarray
        F0 contour.
    sp : ndarray
        Spectral envelope.
    ap : ndarray
        Aperiodicity.
    """
    _f0, t = dio(x, fs, frame_period=frame_period)
    f0 = stonemask(x, _f0, t, fs)
    sp = cheaptrick(x, f0, t, fs, fft_size=fft_size)
    ap = d4c(x, f0, t, fs, order=ap_depth, fft_size=fft_size)
    return f0, sp, ap
                                                                393,21        Bot
```

Merlin-WORLD Acoustic Features

- MGC (Mel-generalized cepstral)
  - 提取到的MFCC特征降低到60维度，方便神经网络训练
- BAP (Band Aperiodicity)
  - 对于48KHz采样，降低到5维
- LF0
  - 基频，1维

Resynth


hms
17.6
17.8
18.0
18.2
18.4
18.6
18.8
19.0
19.2
19.4
19.6
19.8
20.0
20.2
20.4
20.6
20.8
21.0
21.2
21.4
21.6
21.8
22.0
22.2
22.4
dB
-3
-9
-15
-∞
-15
-9
-3
+0 dB
Hz
21k
20k
19k
18k
17k
16k
15k
14k
13k
12k
11k
10k
9k
8k
7k
6k
5k
4k
3k
2k
1k
0:51.250

参考资料

- 论文PDF：https://github.com/tuanad121/Python-WORLD/tree/master/doc/lit
- 科普博文：https://me.csdn.net/xmdxcsj
- Merlin工具脚本： https://github.com/CSTR-Edinburgh/merlin/tree/master/misc/scripts/vocoder/world
- Merlin中文手册：https://mtts.readthedocs.io/zh_CN/latest/merlin.html#merlin-vocoder
- Tacotron整合：https://github.com/Rayhane-mamah/Tacotron-2/issues/304
- 标贝开源中文语音合成数据下载：http://www.data-baker.com/open_source.html