1. 可以从语音质量，合成速度以及模型复杂度多个角度去入手
2. 语音能不能实现一个完全端到端的，直接输出波形，而不是中间会有谱
3. 语音合成的自动评价指标
4. 针对句子级别做的较多，能不能有针对篇章的理解。
5. 语音合成前端的韵律部分。
6. 扩散模型和语音合成