**WAVE-TACOTRON: SPECTROGRAM-FREE END-TO-END TEXT-TO-SPEECH SYNTHESIS**

|  |  |
| --- | --- |
| **Abstract**  Mô tả mạng nơ ron sequence-to-sequence  *Input: Văn bản*  *Output: sóng tiếng nói (speech waveform)*  Kiến trúc mở rộng mô hình Tacotron bằng cách kết hợp luồng chuẩn hóa (normalizing flow) vào vòng giải mã tự hồi quy (autoregressive decoder loop).  Sóng đầu ra được mô hình như một chuỗi các block có độ dài cố định không chồng chéo, mỗi block gồm hàng trăm sample.  Sự phụ thuộc lẫn nhau của các mẫu dạng sóng trong mỗi khối được mô hình hóa bằng cách sử dụng luồng chuẩn hóa, cho phép đào tạo và tổng hợp song song. Các phụ thuộc dài hạn hơn được xử lý một cách tự động hồi quy bằng cách điều chỉnh từng luồng trên các khối trước đó.  Mô hình có thể được **tối ưu bằng cực đại likelihood**, ***không cần sử dụng tính năng trung gian***, đặc trưng thiết kế thủ công hoặc các dạng loss bổ sung.  Các hệ thống SOTA Text-to-speech (TTS) ngày nay sử dụng một loạt một loạt các mô hình học riêng biệt: một trong số đó (Tacotron) tạo ra các **đặc trưng trung gian** (spectrograms) từ văn bản, theo sau đó là **bộ phát âm (vocoder) như** WaveRNN (tạo mẫu dạng sóng từ đặc trưng trung gian).  Hệ thống được đề xuất ngược lại **không dùng đại diện trung gian cố định** và học tất cả tham số end2end. Thực nghiệm thể hiện mô hình đề xuất tạo âm thanh chất lượng tiếp cận các hệ thống TTS SOTA với tốc độ được cải thiện đáng kể.   1. INTRODUCTION |  |