Giới thiệu về cách đánh giá

Để đưa ra cách đánh giá khách quan nhất về chất lượng giọng nói của mô hình, chúng tôi đưa ra 4 tiêu chí sau khi nhận xét về 1 đoạn audio được tổng hợp từ văn bản, đó là:

- (1) chất lượng về **âm thanh**: (âm thanh có rõ không, cò rè không?)
- (2) chất lượng về **phát âm từng từ**: (cách phát âm có tốt không?)
- (3) chất lượng về **độ ngắt nghỉ**: (có ngắt nghĩ giữa những nhịp ngắt không, ví dụ về dấu câu,…)
- (4) chất lượng về **đánh giá chung**: (có dùng đúng các từ chuyên ngành không, có lên xuống giữa những câu hỏi, câu cảm thán, giọng nói có truyền cảm không,…)

Với 4 tiêu chí đánh giá này, chúng tôi cho thang điểm từ 1 đến 5, sau đó lấy trung bình cộng để được điểm cá nhân (R) của đoạn text đó. Tương tự ta lấy trung bình cộng của điểm cá nhân (R) này sẽ cho ra điểm số MOS của 1 người nghe. Cuối cùng chúng tôi lấy trung bình cộng của hai người nghe để cho ra kết quả MOS cuối cùng của hệ thống.

Kết quả thực nghiệm

Các đoạn văn bản trên khi xuất ra file audio có độ dài từ 1 đến 11 giây. Và đánh là kết quả điểm số MOS của chúng tôi khi đánh giá mô hình Tactron2 và WaveGlow này.

System	MOS
Tacotron2 + WaveGlow (người thứ 1)	3.82
Tacotron2 + WaveGlow (người thứ 2)	3.65
Tacotron2 + WaveGlow (trung bình)	**3.735**

Bảng 5: Điểm số ý kiến trung bình (MOS)

Như vậy, từ những đánh giá khách quan bằng cảm nhận của tai nghe thông qua 4 tiêu chí đánh giá, chúng tôi đã đưa ra số điểm MOS cho mô hình tổng hợp giọng nói của chúng tôi là 3.735.