

7.1. Evaluarea performanței sistemelor vocale

7.1.1. ASR: Word Error Rate (WER)

Metrica standard pentru evaluarea sistemelor ASR este Word Error Rate (WER). Aceasta măsoară discrepanța dintre transcrierea de referință (ground truth) și cea generată de sistem (ipoteza). WER se calculează pe baza distanței Levenshtein la nivel de cuvânt, care este numărul minim de operații de substituție (S), ștergere (D) și inserție (I) necesare pentru a transforma ipoteza în referință.

Formula este:

$$WER = \frac{S + D + I}{N}$$

unde N este numărul total de cuvinte din transcrierea de referință.

- Substituție (S): Un cuvânt din referință este înlocuit cu un alt cuvânt în ipoteză (ex: "masă" -> "casă").
- Ștergere (D): Un cuvânt din referință lipsește din ipoteză (ex: "masă mare" -> "masă").
- Inserție (I): Un cuvânt apare în ipoteză, dar nu și în referință (ex: "masă" -> "o masă").

Un WER mai mic indică o performanță mai bună. Deși este standardul industriei, WER are limitări: tratează toate erorile în mod egal, fără a ține cont de impactul lor semantic (ex: a confunda "nu" cu "da" este mult mai grav decât a confunda "o" cu "un", dar ambele sunt contorizate ca o singură substituție).

7.1.2. TTS: Mean Opinion Score (MOS)

Evaluarea calității sistemelor TTS este fundamental subiectivă, deoarece scopul este de a genera o vorbire care sună natural și plăcut pentru un ascultător uman. Metrica "gold standard" este Mean Opinion Score (MOS).

Într-un test MOS, un grup de evaluatori umani ascultă mostre de vorbire sintetizată și le acordă o notă pe o scară de la 1 la 5, corespunzătoare calității percepute :

- 5: Excelent (imperceptibil diferit de vorbirea umană)
- 4: Bun
- 3: Satisfăcător
- 2: Slab
- 1: Nesatisfăcător

Scorul final pentru un sistem este media notelor acordate de toți evaluatorii. Deși este cea mai fiabilă măsură a calității perceptive, realizarea testelor MOS este un proces costisitor, consumator de timp și dificil de standardizat. Din acest motiv, se folosesc și metrice obiective (automate), precum Mel-Cepstral Distortion (MCD), care măsoară distanța spectrală dintre semnalul sintetizat și cel de referință, deși corelația acestora cu percepția umană nu este perfectă.