Damiana Buono

Numero matricola: 0522501592

PIPELINE METODOLOGICA PER L'ANALISI SEMANTICA E LA VALUTAZIONE DEGLI ERRORI DI PRONUNCIA NELLA LINGUA THAI

Relatore: Loredana Caruccio
Correlatore: Dott. Bernardo Breve

Corso di Laurea Magistrale
in Informatica

Università Degli Studi
DI Salerno

Damiana Buono
Matricola N.
0522501592

01. Introduzione al Problema

Crescente diffusione di **tecnologie vocali**

Progresso nei modelli **ASR (Automatic Speech Recognition)**

Le **lingue tonali** (es. Thai) pongono sfide specifiche:

- Dipendenza semantica da toni e durate vocaliche
- Difficoltà di riconoscimento e segmentazione

I modelli ASR multilingue generalisti trascurano aspetti **prosodici** e **fonetici** cruciali

02. Sfide dell’ASR per la lingua thailandese

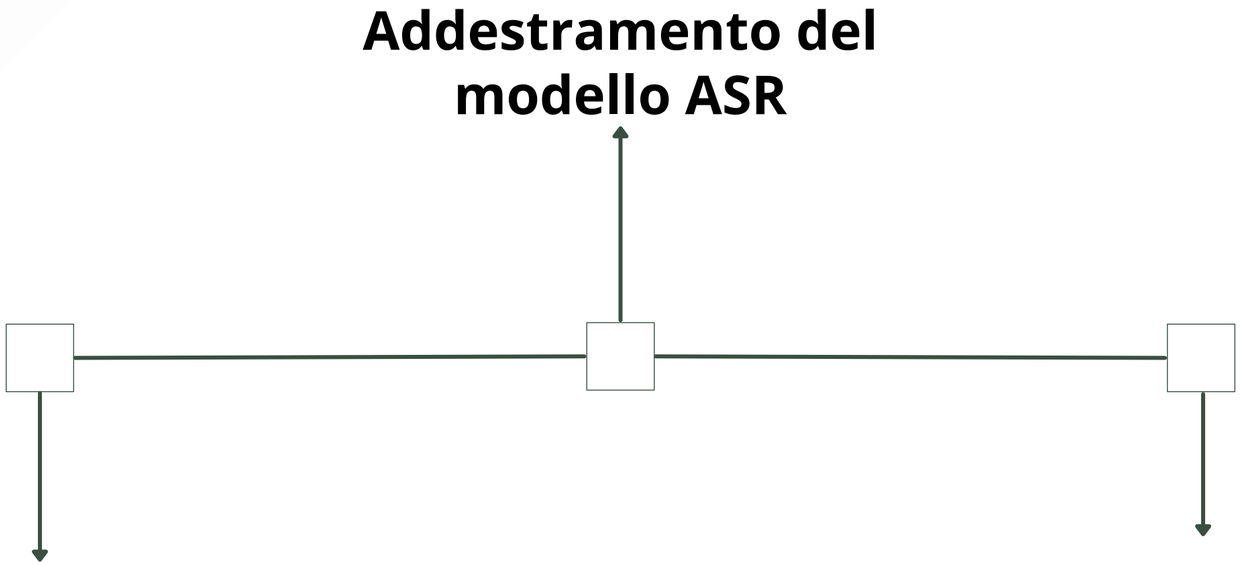
Damiana Buono
Matricola N.
0522501592

Sfide specifiche della lingua thai

Limitazioni dei modelli ASR e dei Corpus disponibili

02. Obiettivi Proposti

Damiana Buono
Matricola N.
0522501592



Damiana Buono
Matricola N.
0522501592

03. Fine-Tuning di un modello ASR

Step 1: Trascrizione automatica (ASR)

Output: trascrizione corretta e pulita del parlato originale

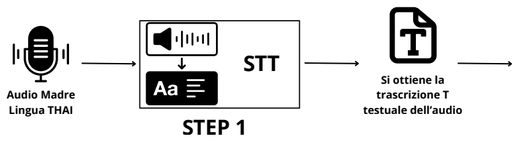
- Esempio trascrizione: **พื้น ทำ ด้วย ไม้ศาล หรือ ฟากสัพ**

Modello ASR: airesearch/wav2vec2-large-xlsr-53-th

- Basato su **Wav2Vec2**
- Ottimizzato per la **lingua thailandese**
- Equilibrio tra **accuratezza e fedeltà**
- **Corpus** di riferimento: *LOTUS*

Segmentazione della trascrizione con **tokenizer newmm:**

- Suddivisione in unità linguistiche coerenti
- Passaggio fondamentale per la successiva fase di injection di errori



Damiana Buono
Matricola N.
0522501592

03. Fine-Tuning di un modello ASR

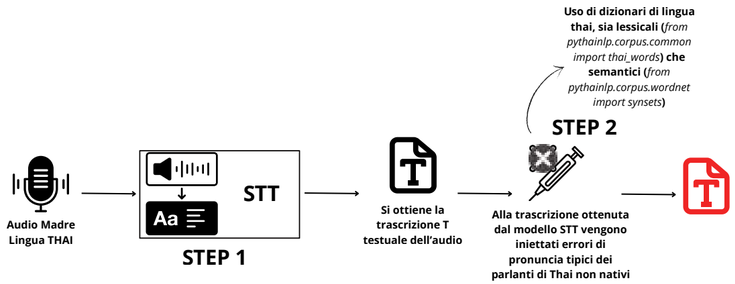
Step 2: Injection controllata di errori fonetici

- **Scopo**: simulare **errori di pronuncia** tipici di studenti Thai
- **Input**: sequenza **tokenizzata** della trascrizione
- **Funzione principale**: *maybe_inject_pronunciation_multi_scaled*
- **Tipologie di errori fonetici**:
  - Consonanti iniziali/finali
  - Vocali lunghe/corte
  - Toni

Output: trascrizione con injection di errori

- Esempio trascrizione: **พื้น ทำ ด้วย ไม้ ศร หรือ ฟาก สัพ**

- **Generazione**: varianti fonetiche plausibili filtrate con dizionario e *WordNet Thai*



Damiana Buono
Matricola N.
0522501592

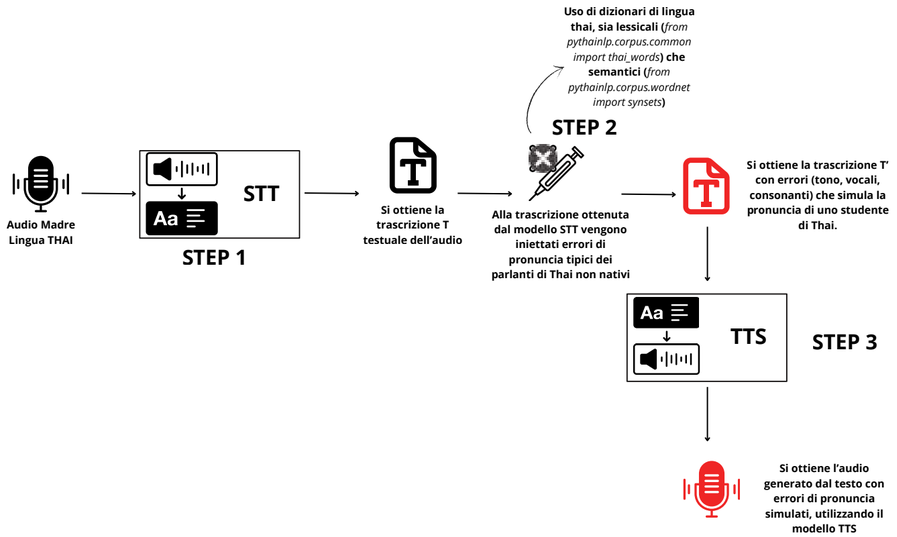
03. Fine-Tuning di un modello ASR

Step 3: Rigenerazione acustica (TTS)

Output: segnale vocale sintetico con **errori di pronuncia realistici**

- **Obiettivo:** creare l'audio della trascrizione alterata
- **Modello TTS:** *khanomtan* (ottimizzato per il thai)

- **Caratteristiche de modello:**
  - Riproduce fedelmente l'input testuale
  - Evita eccessiva normalizzazione automatica
  - Mantiene alterazioni fonetiche e prosodiche introdotte



Damiana Buono
Matricola N.
0522501592

03. Fine-Tuning di un modello ASR

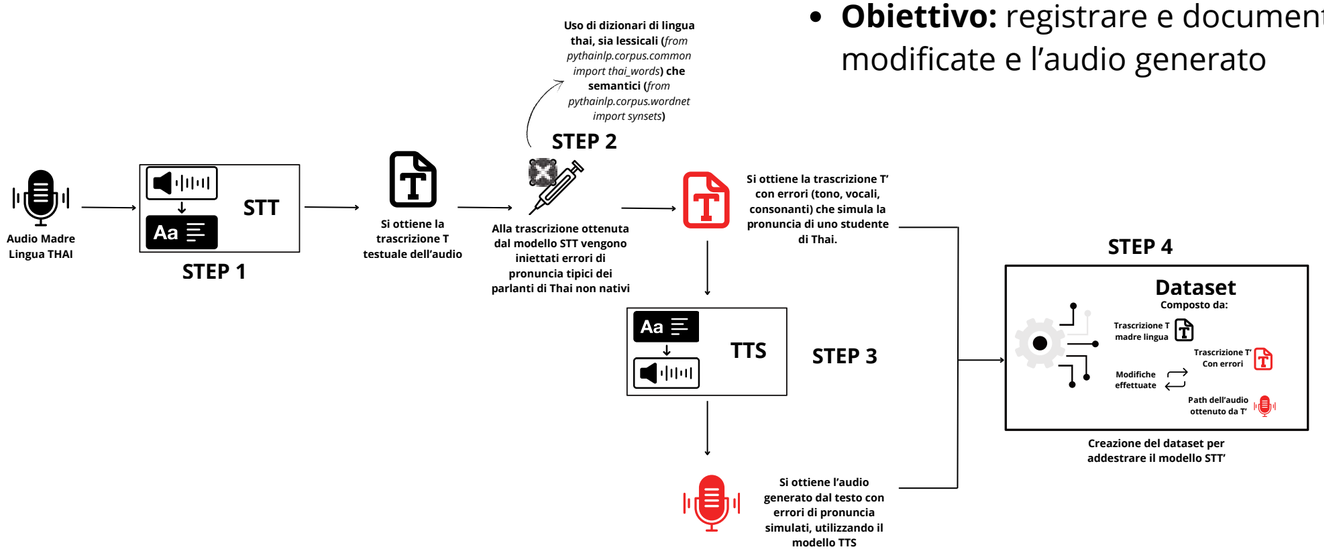
Step 4: Archiviazione dei dati sintetici.

Output: dataset strutturato con:

- Trascrizione originale: **พื้น ทำ ด้วย ไม้ศาล หรือ ฟากสัพ**
- Trascrizione con errori: **พื้น ทำ ด้วย ไม๊ ศร หรือ ฟาก สัพ**
- Dettaglio_modifica: Aggiunto tono: '' in 'ไม๊' | Rimosso ้'' da 'ไม้' | SUONO_VOCALE - 'ศาล': 'า'→'ะ' | SUONO_CONSONANTE - 'ศาล': 'ล'→'ร' | Rimosso 'ะ' da 'ศาล'
- Path file audio **corrotto ottenuto dal modello TTS**

- **Obiettivo:** registrare e documentare le frasi modificate e l'audio generato

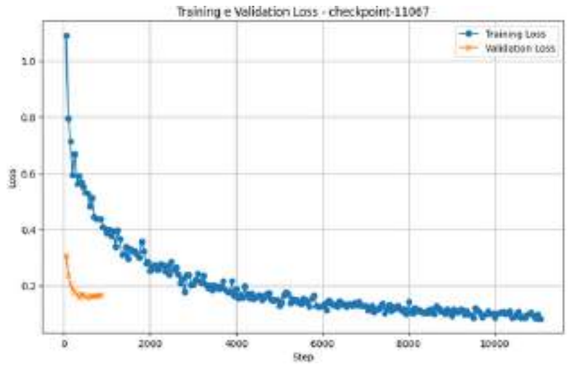
- **Risultato:** dataset sintetico **(3.255 frasi)** utile per:
- Analisi semantica e fonetica
- Addestramento supervisionato di modelli ASR



Damiana Buono
Matricola N.
0522501592

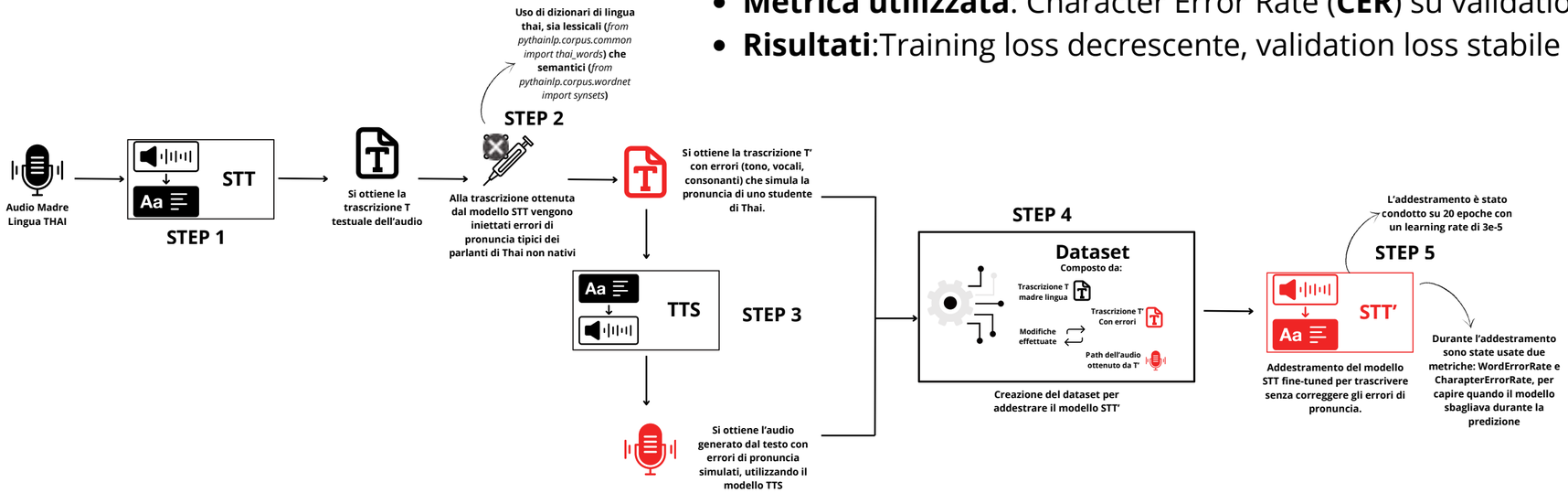
03. Fine-tuning del modello

Step5: Addestramento del modello fine-tuned.



Preparazione dei dati:

- **Modello di utilizzato**: *airesearch/wav2vec2-large-xlsr-53-th*
- **Preprocessing**: Allineamento tra **audio** e **trascrizioni**
- Suddivisione e bilanciamento:
  - Split del dataset: 80% **training** – 10% **validation** – 10% **test**
  - Criteri di bilanciamento:
    - Distribuzione uniforme di errori su: Consonanti iniziali/finali; Vocali corte/lunghe; Toni.
- **Addestramento** : *Learning rate*: 3×10^{-5}; *Max epochs*: 20;
- **Metrica utilizzata**: Character Error Rate (**CER**) su validation set
- **Risultati**:Training loss decrescente, validation loss stabile → un overfitting minimo



04. Scenario Applicativo

Damiana Buono
Matricola N.
0522501592

Caso di Studio: Analisi Semantica delle Produzioni Linguistiche in Thai

Contesto:

- Studenti e utenti non madrelingua
- Livelli di competenza: principiante, intermedio.
- Necessità di un apprendimento naturale e personalizzato

Obiettivo:

- Promuovere un apprendimento linguistico personalizzato
- Valutare in modo integrato pronuncia e il significato
- Fornire un feedback immediato e mirato sul parlato dello studente

Approccio Tecnologico

- Integrazione di modelli ASR avanzati e analisi semantica per comprendere il parlato degli studenti
- Supporto a un apprendimento interattivo centrato sullo studente

04. Pipeline per la valutazione del parlato

Damiana Buono
Matricola N.
0522501592

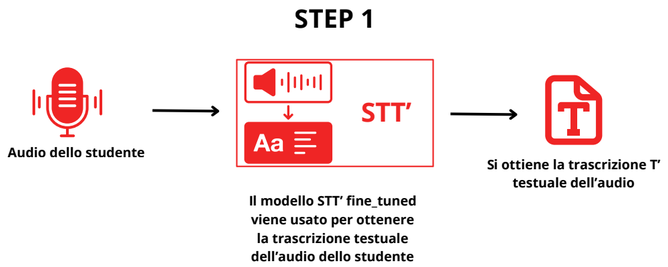
Step 1: Trascrizione Automatica del Parlato (ASR)

Output: trascrizione fedele, senza correzioni automatiche.

- Esempio:
  - Riferimento: ส้วม ทาน ด้อม ใต บรูก เข้าเจ้า อาย หงัก
  - Predizione del modello: ส้วม ทาน ด้อม ใต บรูก เข้าเจ้า อาย หงัก

Obiettivo: preservare gli errori per analisi successive.

Modello: *wav2vec2* fine-tuned su parlato con errori injection di errori



04. Pipeline per la valutazione del parlato

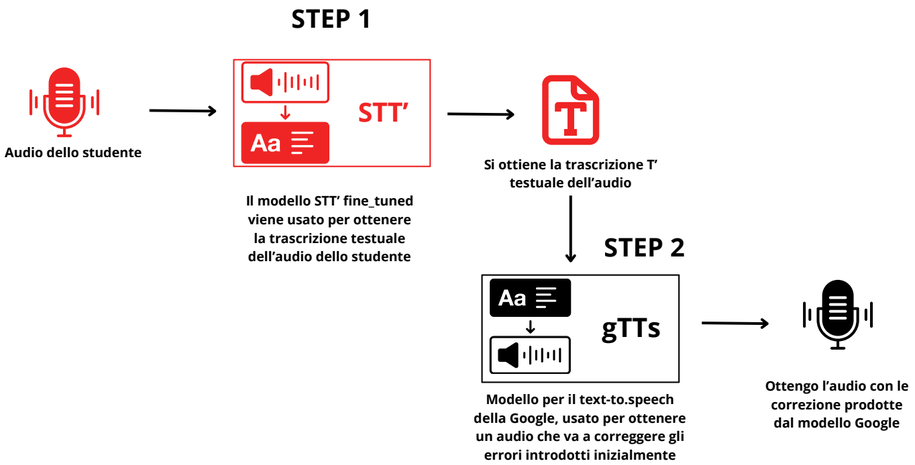
Damiana Buono
Matricola N.
0522501592

Step 2: Generazione dell’Audio di Riferimento (TTS)

Fornisce un riferimento acustico per confronto fonetico e tonale

Crea una **pronuncia corretta standard** a partire dal testo. Audio chiaro, coerente in tono e durata

Strumento: gTTS (Thai)



04. Pipeline per la valutazione del parlato

Damiana Buono
Matricola N.
0522501592

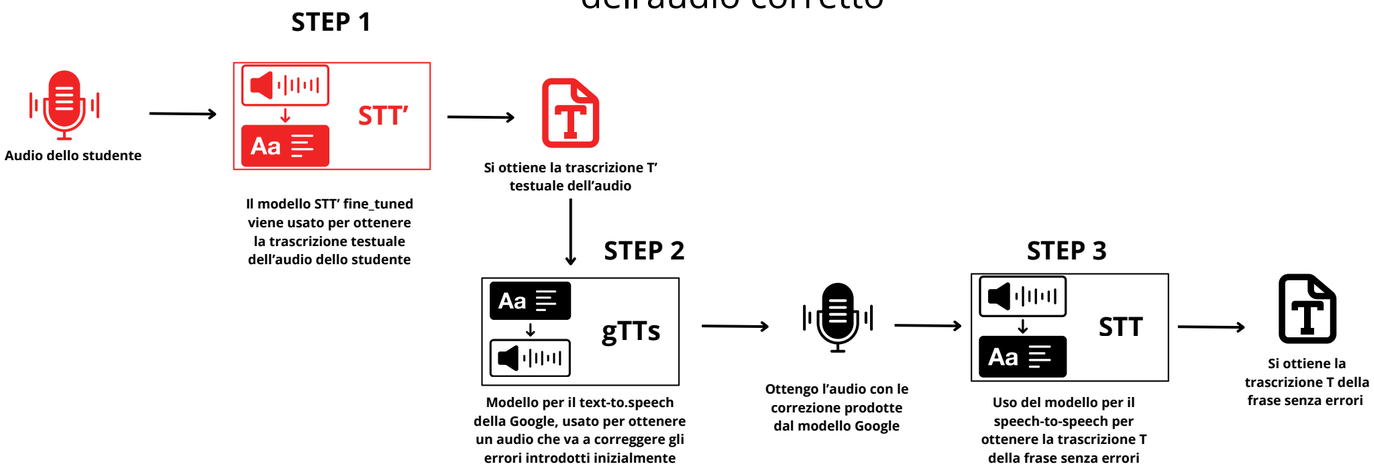
Step 3: Trascrizione dell’Audio di Riferimento (ASR di Riferimento)

Esempio:

- testo ottenuto dalle correzioni inserite da gTTS:
  ส้วม ทาน ดอม เต้ บรู้ เข้า เจ้า อาย หงัก

Produce la versione testuale dell’audio corretto

Modello: wav2vec2-large-xlsr-53-th classico



04. Pipeline per la valutazione del parlato

Damiana Buono
Matricola N.
0522501592

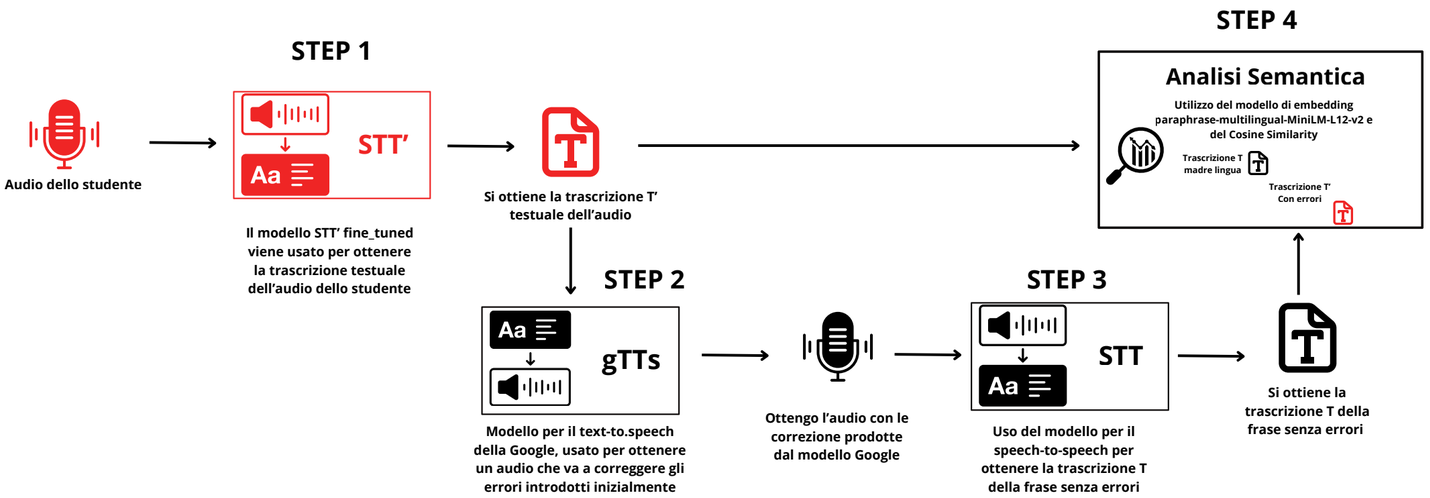
Step 4: Analisi Semantica

Esempio:

- trascrizione pronunciata dallo studente: ส้วม ทาน ด้อม ใต บรูก เข้าเจ้า อาย หงัก
- trascrizione con la corretta pronuncia: ส้วม ทาน ดอม เต้ บรู้ เข้า เจ้า อาย หงัก
- similarità semantica:

Modello: *paraphrase-multilingual-MiniLM-L12-v2*

- **Confronto** tra trascrizione studente ↔ versione corretta
- Calcolo della **similarità semantica** (cosine similarity)



04. Pipeline per la valutazione del parlato

Damiana Buono
Matricola N.
0522501592

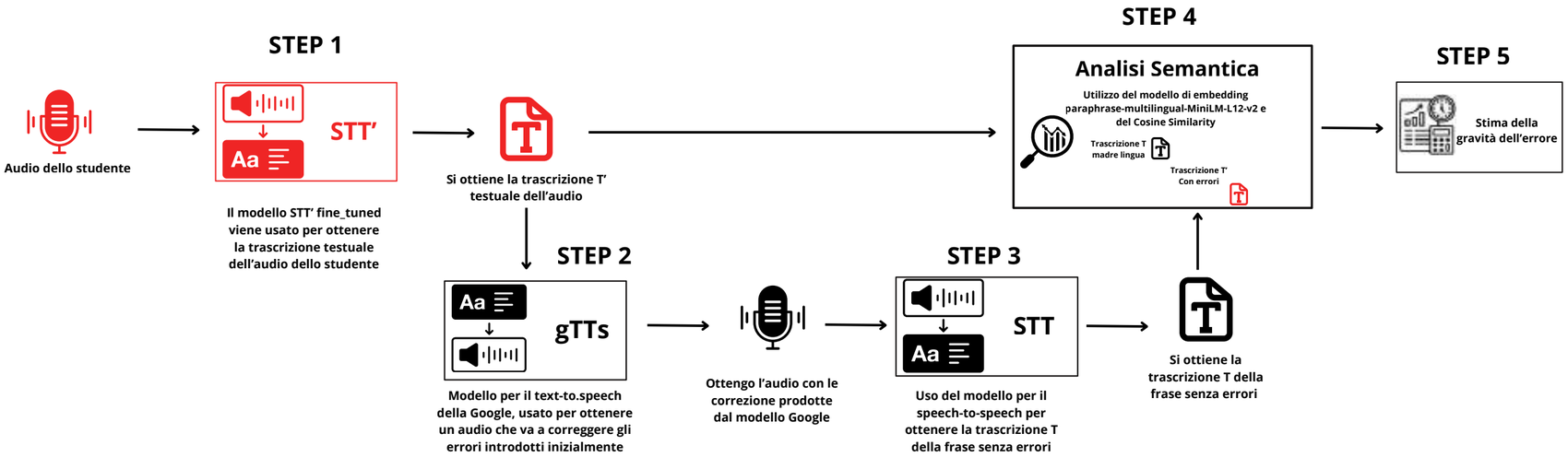
Step 5: Analisi Fonetica e Classificazione della Gravità

Stima della Gravità basata sul **numero di errori inseriti**.

- 0 = Nessuno
- 1 - 3 = Lieve
- 4 - 6 = Medio
- > 6 = Grave

Obiettivo
Identificare e quantificare gli errori fonetici e tonali per **stimare** la gravità complessiva della pronuncia

Analisi **fonetica e tonale,**
Confronto tra **frase studente** e **frase corretta**



04. Pipeline Metodologica

Damiana Buono
Matricola N.
0522501592

Step 6: Generazione del Feedback Linguistico

Il sistema genera un **messaggio descrittivo**, che indica:

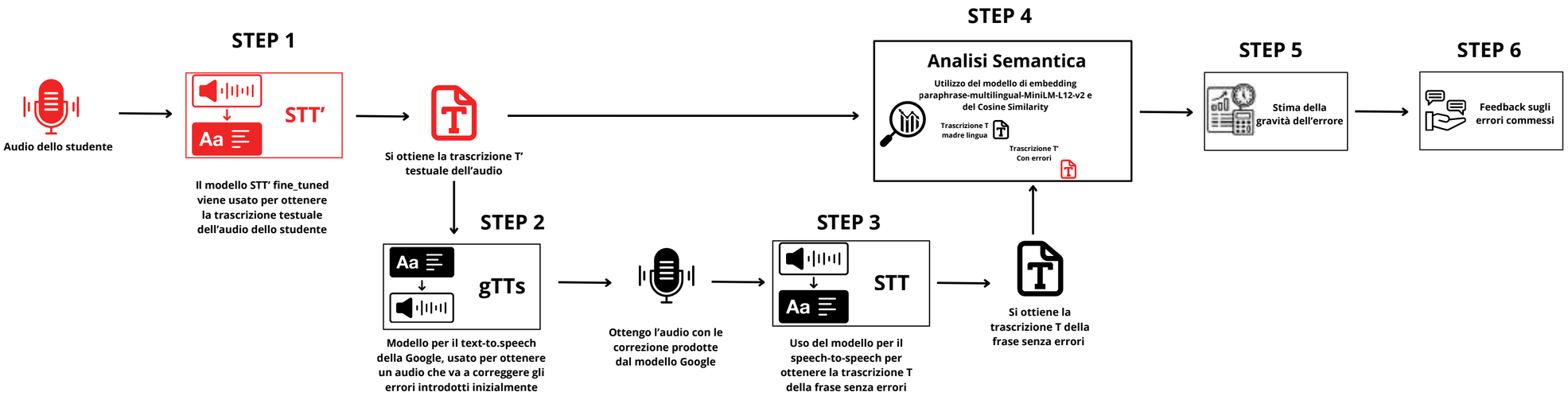
- Dove si trovano gli errori
- Quale tipo di errore è stato commesso (tono, vocale, consonante)
- Come migliorare la pronuncia o la struttura semantica

Obiettivo

Unire i risultati dell'analisi semantica e fonetica per creare un feedback completo, personalizzato e operativo per lo studente.

Esempio feedback:

La frase '**ส้วม ทาน ด้อม ใต บรูก เข้าเจ้า อาย หงัก**' è molto simile a '**ส้วม ทาน ดอม เต้ บรู้ เข้า เจ้า อาย หงัก**' (0.97). Ottimo lavoro!
Solo piccole variazioni fonetiche potrebbero essere migliorate.
Suggerimenti:
- La parola 'ด้อม' è diversa da 'ดอม', ma il significato resta simile
- Migliora la pronuncia di 'ใต', dovrebbe essere 'เต้'
- La parola 'รู' è diversa da 'รู้', ma il significato resta simile
- Migliora la pronuncia di 'ก', dovrebbe essere 'เข้า'
- Migliora la pronuncia di 'เข้าเจ้า', dovrebbe essere 'เจ้า'



05. Domande di ricerca (Research Questions)

Damiana Buono
Matricola N.
0522501592

RQ2: Quanto impatta la quantità di errori rispetto alla capacità di trascrizione fedele?

RQ1: Quanto impatta la tipologia di errore rispetto alla capacità di trascrizione fedele?

RQ3: Quanto il modello di valutazione degli errori altera il significato semantico della frase rispetto alla versione corretta?

05. Metriche di valutazione

Damiana Buono
Matricola N.
0522501592

RQ1 & RQ2
Accuratezza fonetica**:**
Character Error Rate (**CER**), ovvero percentuale di caratteri trascritti in modo errato dal modello di riconoscimento del testo.

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

RQ3
Coerenza semantica**:**
Cosine Similarity (basata su MiniLM-L12-v2)

$$CosSim(A, B) = \frac{A \cdot B}{\|A\|\|B\|}$$

Stima della Gravità basata sul **numero di errori inseriti**.

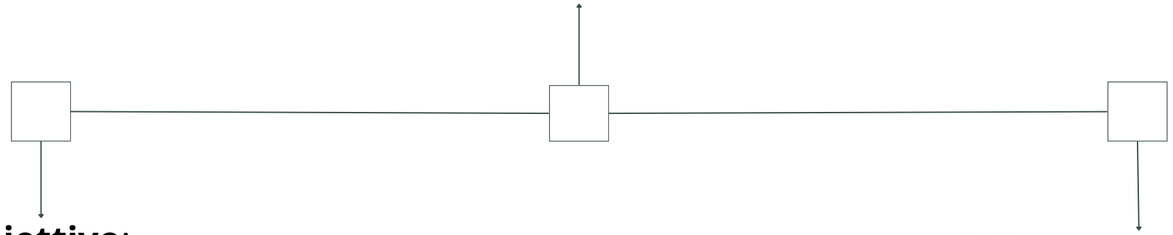
- 0 = Nessuno
- 1 - 3 = Lieve
- 4 - 6 = Medio
- > 6 = Grave

05. Configurazione sperimentale

Damiana Buono
Matricola N.
0522501592

Dataset:

- Basato su **LOTUS** corpus
- Versione estesa con injection controllato di errori del parlato
- Uso solo dei dati di **test**
- **326** frasi nel **test set**



Obiettivo:
Analizzare l'impatto delle tipologie e quantità di errori fonetici sulla trascrizione e sul significato delle frasi.

Modello:

- *wav2vec2-large-xlsr-53-th fine-tuned*
- Denominato *Eval Model*

Damiana Buono
Matricola N.
0522501592

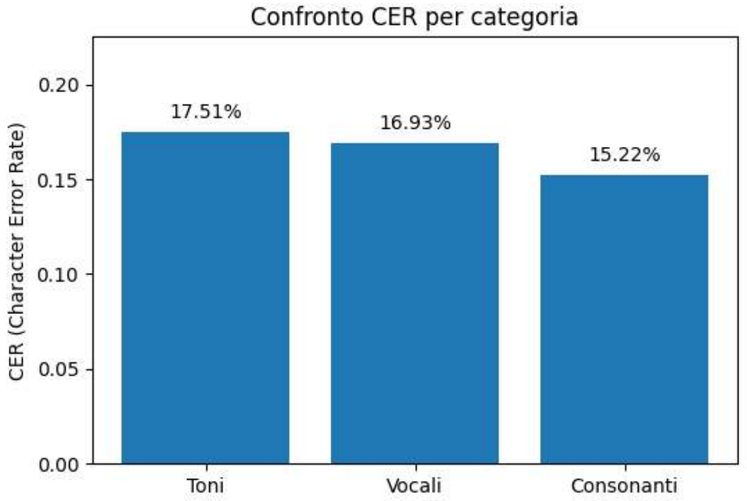
06. Analisi dei Risultati

RQ1: Quanto influisce il tipo di errore fonetico sull'accuratezza della trascrizione?

Categoria	CER medio	W Shapiro-Wilk	p-value
Toni	0.17	0.95	0.67
Vocali	0.16	0.93	0.54
Consonanti	0.15	0.92	0.42

Analisi condotte

- Calcolo del **Character Error Rate (CER)** per tre tipologie di errore: toni, vocali e consonanti.
- **Test di Shapiro–Wilk** → verifica della normalità dei dati (p > 0.05 → distribuzione normale).
- Applicazione di **test ANOVA** per confrontare le medie dei gruppi.
- Dataset bilanciato: 10 frasi per ogni categoria di errore, stessa complessità sintattica.



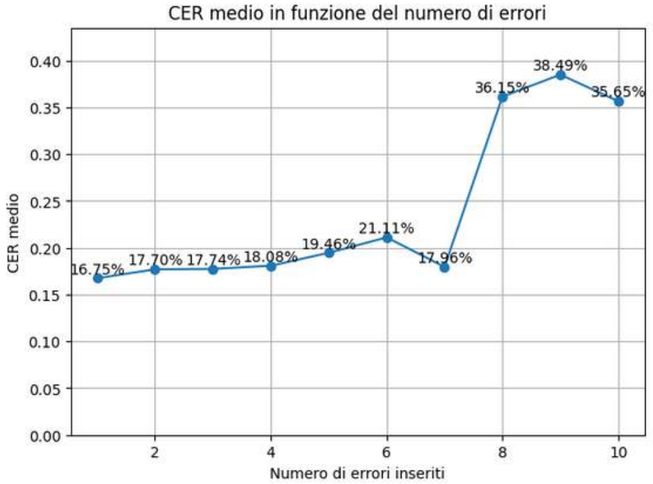
Risultati ottenuti:

- **Shapiro–Wilk**: p > 0.05 → dati normalmente distribuiti.
- **ANOVA**: p = 0.37 → nessuna differenza significativa.
- **Conclusione**: il tipo di errore non influisce significativamente sulla capacità di trascrizione.
- Il **modello** mostra **robustezza** rispetto alle variazioni fonetiche.

Damiana Buono
Matricola N.
0522501592

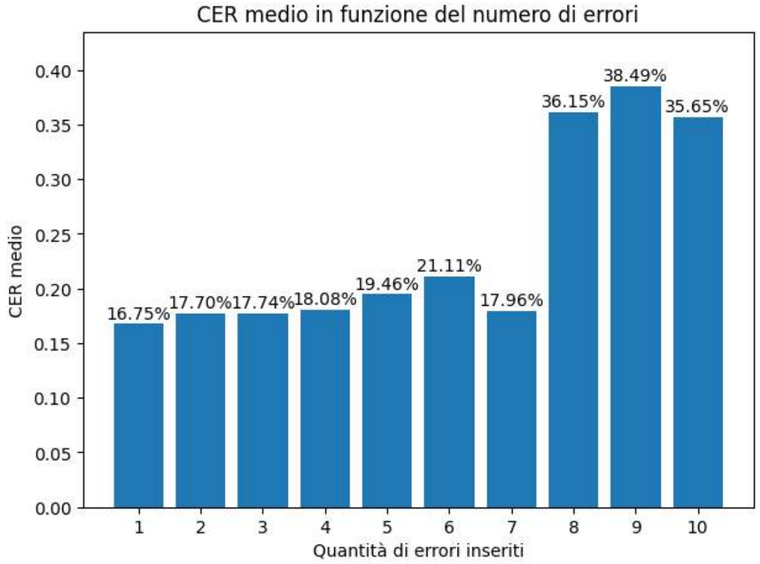
06. Analisi dei Risultati

RQ2: Quanto impatta la quantità di errori rispetto alla capacità di trascrizione fedele?



Analisi condotte

- Calcolo del **CER medio** per frasi con da **1 a 10 errori** fonetici.
- Test di **Shapiro–Wilk** sulla distribuzione del CER → verifica normalità.
- In base al risultato:
- se normale → correlazione Pearson;
- se non normale → correlazione Spearman.



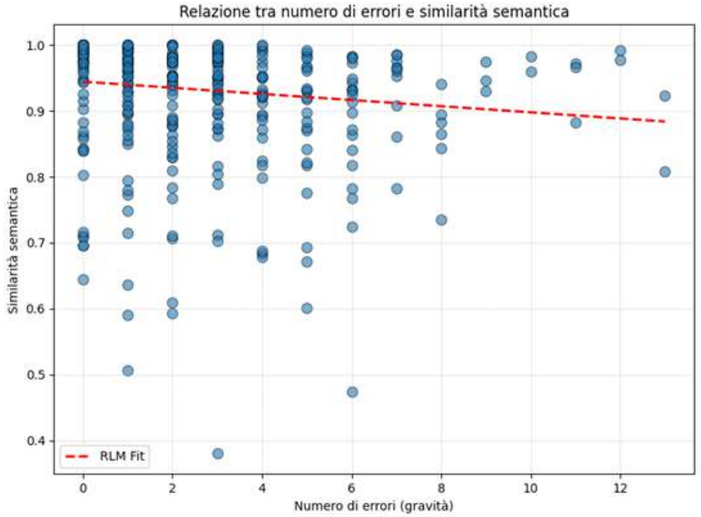
Risultati ottenuti

- **Shapiro–Wilk:** tutti i valori dei gruppi sono normali, solo il sesto gruppo ha p=0.036; p< 0.05 → dati non normali.
- Applicata **correlazione di Spearman**: ρ = 0.89, p = 0.001 → correlazione positiva forte.
- CER stabile fino a 7 errori, poi crescita esponenziale.
- Conclusione: il **modello** mantiene fedeltà fino a un livello medio di errore, ma crolla oltre la soglia critica.

Damiana Buono
Matricola N.
0522501592

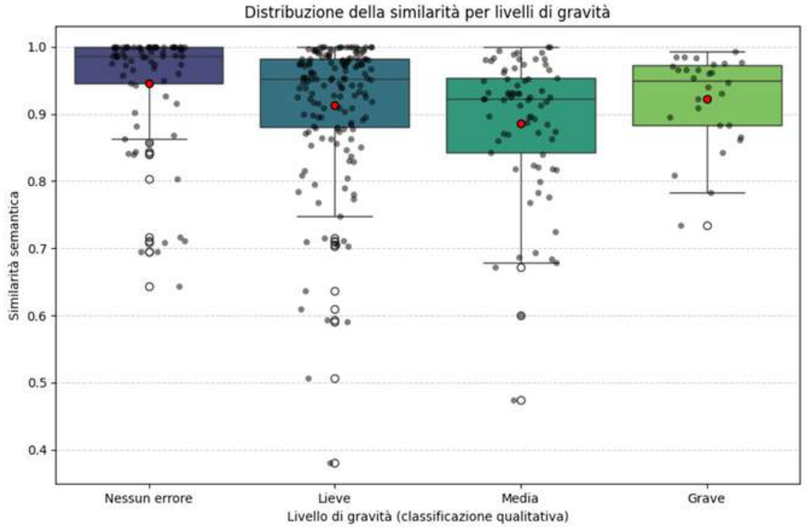
06. Analisi dei Risultati

RQ3: Quanto l' Eval Model altera il significato semantico della frase rispetto alla versione corretta?



Analisi condotte

- Definizione dei **livelli di gravità** basati sul numero di errori: 0 = Nessuno; 1–3 = Lieve; 4–6 = Medio; > 6 Grave
- Calcolo della **cosine similarity** tra frase errata e corretta tramite MiniLM-L12-v2.
- **Test di Shapiro–Wilk** → verifica normalità ($p < 0.05$ → non normale).
- Applicazione di **Kruskal–Wallis** per differenze tra gruppi.
- Calcolo di **correlazione Spearman** (ρ) e regressione lineare per stimare l'effetto della gravità.



Risultati ottenuti

- **Shapiro–Wilk**: $p < 0.05$ → distribuzione non normale.
- **Kruskal–Wallis**: $p < 0.001$ → differenze significative tra i livelli di gravità.
- **Spearman** $\rho = -0.2977$ ($p < 0.001$) → relazione negativa.
- **Regressione**: $\beta = -0.0047$ → conferma la tendenza decrescente.
- Conclusione: all'aumentare della gravità, la **similarità semantica diminuisce**.
- Tuttavia, nelle frasi con errori lievi o medi, il modello preserva il significato generale.

Damiana Buono
Matricola N.
0522501592

07. Conclusioni

Pipeline proposta per trascrizione fedele del parlato con errori di pronuncia Thai.

Modello **ASR** fine-tuned (*wav2vec2-large-xlsr-53-th*) combinato tokenizzazione e analisi semantica.

Dataset: **3.255** esempi, divisi in training (80%), validation (10%) e test (10%).

Risultati principali:

- Robusto su toni, vocali e consonanti.
- CER stabile fino a ~7 errori per frase.
- Significato preservato per errori lievi/moderati.

Validità della pipeline confermata, ma necessaria estensione a dati reali e valutazioni percettive.

Damiana Buono
Matricola N.
0522501592

07. Lavori Futuri

- **Raccolta** e **integrazione** di dati reali da studenti di Thai .
- Confronto tra **metriche** automatiche e **giudizi umani** per valutare la percezione fonetica.
- Analisi più approfondita di **localizzazione** e **tipologia** degli errori.
- Integrazione di informazioni **prosodiche** (intonazione, durata, frequenza).
- Ottimizzazione computazionale (distillation, quantization) per inferenza efficiente.
- Sviluppo di **strumenti didattici** interattivi per **feedback** personalizzato sulla pronuncia.
- Estensione a dialetti, code-switching e altre lingue tonali.

Damiana Buono

Numero matricola: 0522501592

PIPELINE METODOLOGICA PER L'ANALISI SEMANTICA E LA VALUTAZIONE DEGLI ERRORI DI PRONUNCIA NELLA LINGUA THAI

Grazie per l'attenzione

Relatore: Loredana Caruccio
Correlatore: Dott. Bernardo Breve

Corso di Laurea Magistrale
in Informatica

Università Degli Studi
DI Salerno