

LA TRASCRIZIONE FONETICA AUTOMATIZZATA NEI SISTEMI DI SINTESI DELLA VOCE: UNA NUOVA TECNICA DI APPRENDIMENTO E CARATTERISTICHE DEI LESSICI PER L'ADDESTRAMENTO

Paolo Massimino e Alberto Pacchiotti

Loquendo S.p.A.

RIASSUNTO

Uno dei passi fondamentali nella conversione da testo a voce (TTS: Text-To-Speech) è la trascrizione fonetica, cioè il passaggio da un testo ortografico alla sua equivalente versione in fonemi.

Per molte lingue (ad esempio, quelle romanze), questa conversione può essere ottenuta da un programma che si basa su un certo numero di regole ed una lista di eccezioni limitata.

Queste regole possono essere scritte da un esperto fonetista, e solo le eccezioni e le parole straniere richiedono l'aggiunta di una lista integrativa.

Questo tipo di soluzione diventa invece molto difficile per alcune lingue, ad esempio l'Inglese e l'Olandese, dove sembra che solo dando un elenco (il "lessico fonetico"), dove ogni parola è accompagnata dalla sua trascrizione, si possa ottenere una trascrizione fonetica corretta.

L'utilizzo di un lessico ha comunque i suoi svantaggi: per quanto grande esso sia, difficilmente riuscirà a contenere tutte le parole di una lingua, oltre alle parole straniere, ai nomi propri, toponimi e neologismi.

Inoltre, un lessico fonetico di grandi dimensioni può creare seri problemi se deve essere utilizzato in dispositivi con un limitato spazio di memoria.

Per ovviare a questi problemi, è stato realizzato un nuovo algoritmo, detto "automa fonetico", che ha due funzioni fondamentali: "comprimere" un grande lessico fonetico in un numero limitato di regole (ottenute durante una fase di apprendimento automatico), e "generalizzare", cioè dare la corretta trascrizione fonetica non solo per le parole del lessico medesimo, ma anche per parole del tutto nuove.

Le caratteristiche che devono essere prese in esame nella scelta del lessico fonetico, da utilizzare nella fase di apprendimento, sono varie: le dimensioni, il grado di "copertura" rispetto alle sequenze di fonemi rappresentative della lingua, l'inclusione o meno di allofoni e di fonemi appartenenti ad altre lingue.

L'algoritmo proposto è in grado di determinare l'ampiezza ottimale del contesto ortografico, da considerare durante il processo di estrapolazione delle regole di trascrizione fonetica: ciò permette di generare regole di trascrizione particolarmente ottimizzate in termini di complessità computazionale.

Bibliografia

[1] Mana, Franco / Massimino, Paolo / Pacchiotti, Alberto (2001): "Using machine learning techniques for grapheme to phoneme transcription", In *EUROSPEECH-2001*, 1915-1918.