

CAP.2: Asistenți vocali

Contextul apariției:

Vorbirea este un mijloc versatil de comunicare. Ea este compusă din : mesaje lingvistice (de exemplu, Mesaj și limbă), vorbitor (de exemplu, caracteristicile emoționale și fiziologice ale aparatului vocal) și mediul înconjurător (de exemplu, unde a fost produs și transmis mesajul vocal). Chiar dacă astfel de informații sunt codificate într-o formă complexă, oamenii le pot descifra relativ ușor pe majoritatea.

Această abilitate umană a inspirat cercetătorii să elaboreze sisteme care să imite o astfel de abilitate. De la foneticieni la ingineri, cercetătorii au lucrat pe mai multe fronturi pentru a decoda majoritatea informațiilor din semnalele acustice. Unele fronturi includ sarcini cum ar fi identificarea vorbitorului prin recunoașterea vocii, detectarea limbii vorbite, transcrierea discursului, traducerea discursului și înțelegerea vorbirii.

Recunoașterea vocală este folosită pe scară largă pentru diverse aplicații în diverse gadget-uri și sisteme electronice. Aplicația principală este de a susține căutarea în baza de date cu ajutorul serviciilor IVRS (Sistem integrat de răspuns vocal). Aplicația IVRS îi ajută pe utilizatori să caute informațiile fie prin selectarea tactilă a meniurilor, fie prin intermediul comenzilor vocale.

Recunoașterea automată a vorbirii (ASR) a fost în centrul atenției pentru mulți cercetători de mai multe decenii. În acest tip de recunoaștere, mesajul lingvistic este informația de interes. Aplicațiile de recunoaștere a vorbirii variază de la dictarea unui text la generarea de subtitrări în timp real pentru o emisiune de televiziune. În ciuda capacității umane, cercetătorii au aflat că extragerea informațiilor din discurs nu este un proces simplu. Variabilitatea în vorbire datorită limbajului, factorilor fiziologici și de mediu îi provoacă pe cercetători să extragă corect informațiile relevante din semnalul de vorbire. În ciuda tuturor provocărilor, cercetătorii au făcut progrese semnificative în tehnologie, astfel încât este posibil să se dezvolte aplicații cu funcție de vorbire.

Problema recunoașterii vorbirii poate fi descrisă ca o funcție, care pornește de la dovezile acustice ale unui singur cuvânt sau ale unei succesiune de cuvinte. Fie $X = (x_1, x_2, x_3, \dots, x_t)$ reprezintă dovezile acustice, care sunt generate în timp (indicate de indicele t) dintr-un semnal de vorbire dat și aparțin setului complet de secvențe acustice, χ . Fie $W = (w_1, w_2, w_3, \dots, w_n)$ o secvență de cuvinte n , fiecare aparținând unui set fix și cunoscut de cuvinte posibile, ω . Există două cadre pentru a descrie funcția de recunoaștere a vorbirii: șablon și statistică.