

Spektro-temporális felbontáson alapuló elemzési módszerek a beszédfelismerésben

Kovács György

V. Programtervező-matematikus

Témavezető: Tóth László tudományos főmunkatárs

SZTE TTIK Mesterséges Intelligencia kutatócsoport

A beszédfelismerésben alkalmazott jellemzőkinyerés hagyományos módszerei matematikai alapokra épülnek, csupán a legalapvetőbb tulajdonságait figyelembe véve az artikulációnak és hallásnak. Habár nem feltétlenül szükségszerű, hogy az emberi hallás utánzására törekvő módszerek teljesítménye felülmúlja a tisztán matematikai algoritmusokét, általánosságban ésszerű elvárás, hogy megfelelőbb viselkedést tapasztaljunk azon eljárásoktól, amelyek jobban közelítik az emberi hallás tulajdonságait.

Az egyik ilyen tulajdonság, hogy az agykérgi sejtek receptív mezői egyidejűleg reagálnak a hangjel időbeli és frekvenciabeli (azaz spektro-temporális) változásaira. Összehasonlítva a beszédfelismerésben hagyományosan alkalmazott MFCC (Mel-Frequency Cepstral Coefficient) reprezentációval, utóbbi felbontása jóval keskenyebb idő-, és szélesebb frekvenciatartományban. Ezzel összhangban, Jake Bouvrie és Tony Ezzat korábbi munkája nyomán, az MFCC algoritmus egyszerű módosítását alkalmaztam, amely a hagyományos MFCC technika keskeny idő-, és globális frekvenciatartományú ablakaival szemben, lokalizált spektro-temporális ablakokat használ a spektrális reprezentáción. Célom megmutatni, hogy a korábbi munkák jellemzőkinyerési idejét jelentősen csökkentve, a fonémák osztályozása terén a felkínált jellemzők továbbra sem rosszabbak a hagyományos jellemzőkészletnél. Továbbá bizonyos zajokkal szemben robusztusabbak is annál.