

2010. január 22-én

[Zainkó Csaba](#) és [Csapó Tamás Gábor](#),

a BME Távközlési és Médiainformatikai Tanszékének munkatársai bocsátották vitára a címben megfogalmazott kérdést.

Spontán beszéd géppel?

A mai beszédszintetizátorok jelentős része felolvasott beszéden alapul. A szintetizátor típusától függően a szükséges beszédfelvétel hossza néhány órától egészen a 40-50 óráig is terjedhet. A felvételeket többnyire képzett bemondókkal készítjük, ügyelve az egyenletes tempóra, hangmagasságra, stílusra. Az így készült hangadatbázisokat feldolgozzuk részben automatikusan, részben kézzel. A szintetizátorok ezeket a hangadatbázisokat felhasználva felolvasott beszédet állítanak elő alapesetben. Régebben a beszédszintetizátorokkal szemben az volt az elvárás, hogy érthető és jó minőségű beszédet állítsanak elő. Később, amikor ez többé-kevésbé teljesült, igény jelentkezett a természetes beszéd előállítására.

□ □

De milyen a természetes beszéd?

□

Eleinte a robotos hangzás idején azt gondoltuk, hogy ha a hangszínezet emberi lesz, akkor elértük ezt a célt. A hangszínezet ma már közel emberinek mondható, de a természetesség érzete még mindig nem teljesült. Akkor merre tovább?

A beszédszintézis általában nem szereti a spontán beszédet, mivel a spontán beszéd különböző tulajdonságai technológiai akadályként jelennek meg. A felvételek során ezeket kerültük, illetve a feldolgozás során eltávolítottuk a hanganyagból.

□ □ Akkor a spontán beszéd?

A jelenlegi kutatásunkban azt vizsgáljuk, hogy a spontán beszéd és a gépi beszédszintetizátorok közötti ellentét hogyan oldható fel.

Ha a spontán beszéd tulajdonságait átültetjük a szintetizált beszédbe, akkor természetesebb lesz?

Az egyik megoldás az lehetne, hogy spontánbeszéd-felvételekből kiindulva készítünk beszédszintetizátort, de ez a sok éves tapasztalatunk alapján túl nagy lépés lenne elsőre. A másik vizsgált irány az, hogy hogyan hozható létre egy olyan beszédszintetizátor, amely felolvasott beszéden alapul, és mégis spontánabb a hangzása a korábbiaknál. Vizsgáljuk, hogy miként módosíthatóak az eddigi eljárások úgy, hogy a spontán beszéd elemeit, tulajdonságait is figyelembe vegye, illetve modellezni tudja azokat.

A találkozón azt vitattuk meg, hogy a spontán beszéd milyen elemei elfogadhatóak a szintetizált beszédben, illetve van-e létjogosultsága az ilyen gépi beszédnek. A könnyebb értékelhetőség érdekében a mintákat előzetesen egy meghallgatásos teszt keretében megismerhették a résztvevők és az érdeklődők.

Műhelytalálkozó 2010. január 22.

Írta: Administrator

2010. január 06. szerda, 09:48 - Módosítás: 2010. szeptember 10. péntek, 06:38

A találkozón elhangzott beszélgetésről hangfelvétel készült, amelynek tartalmi kivonata [itt](#) olvasható.