FAKULTA APLIKOVANÝCH VĚD ZÁPADOČESKÉ UNIVERZITY V PLZNI

STUDENTSKÁ VĚDECKÁ KONFERENCE 2017

Využití artikulačních příznaků v syntéze řeči

Martin Matura[1]

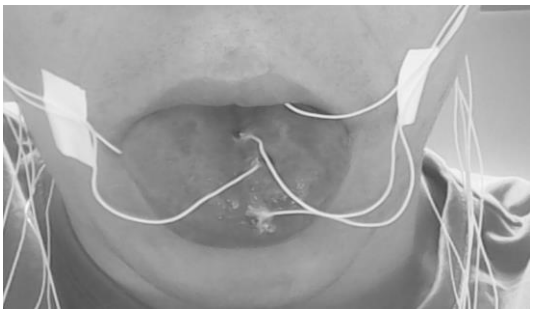
1. Úvod

Pro syntézu řeči v poslední době používáme nejčastěji metodu výběru jednotek. Každá jednotka má definované určité parametry, příznaky, podle kterých se vybírá posloupnost jednotek tak, aby odpovídala požadované promluvě. Mezi příznaky nyní řadíme informaci o frekvenčním spektru v podobě MFCC (z angl. *Mel-frequency cepstrum coeficient*) (M), energii signálu (E) a jeho základní frekvenci (F). Další přídatnou informaci, která by podle Richmond a King (2016) mohla vést k lepší kvalitě syntetizované řeči, by mohly poskytnout artikulační příznaky.

Artikulační příznaky popisují artikulační trajektorie, tj. pohyb, který vykonávají artikulátory (jazyk, rty, spodní čelist, atd.) při vytváření řeči. Trajektorie získáváme zpracováním dat z elektromagnetického artikulografu, který je schopen tyto pohyby měřit pomocí senzorů.

2. Získání dat

K získání artikulačních trajektorií jsme použili 3D elektromagnetický artikulograf AG501. Pro snímání jsme použili 7 senzorů – tři referenční, které snímaly pozici hlavy a čtyři pro měření trajektorií, z nichž tři byly upevněny na jazyku (obr. 1) a jeden na spodní čelisti.



Obrázek 1: Upevnění senzorů na jazyku

K našemu experimentu jsme pak využili 380 nahraných vět (přibližně 35 minut řeči). Ze zvukových nahrávek byla vytvořena databáze jednotek (difonů) a nahraná data jsme museli předzpracovat. Byla provedena filtrace šumu, korekce dat od pohybů hlavy a přiřazení artikulačních trajektorií k odpovídajícím difonům. Takto připravená data jsme použili k syntéze vět a zajímalo nás, zda začleněním artikulačních příznaků vylepšíme výslednou kvalitu syntézy a také, zda bychom těmito příznaky mohli nahradit informaci z MFCC – to by bylo možné, pokud by se prokázala korelace mezi standardní konkatenační cenou (MEF) a konkatenační cenou, kde jsou MFCC nahrazeny artikulačními příznaky (AEF).

[1] student doktorského studijního programu Aplikované vědy a informatika, obor Kybernetika, e-mail: mate221@kky.zcu.cz

3. Korelace s používanými příznaky

Po přidání artikulačních příznaků jsme v kvalitě výsledné syntézy žádné znatelné zlepšení nezaznamenali, což s největší pravděpodobností způsobuje malý počet dat, ze kterých se hlas pro syntézu vytvářel. Dále jsme pak zjišťovali, jaká je korelace mezi konkatenační cenou s příznaky MEF a konkatenační cenou s příznaky AEF podle:

$$r = \frac{\sum_m \sum_n (A_{mn} - \bar{A})(B_{mn} - \bar{B})}{\sqrt{\left(\sum_m \sum_n (A_{mn} - \bar{A})^2\right)\left(\sum_m \sum_n (B_{mn} - \bar{B})^2\right)}}. \quad (1)$$

Korelaci jsme zjišťovali pro různý počet vět, jak ukazuje tabulka 1, která přehledně shrnuje výsledné korelační koeficienty. Jak je vidět, korelační koeficient se u sta vět pohybuje kolem 0.77, což značí poměrně velkou míru korelace.

Počet vět	**1**	**10**	**100**
Korelační koeficient	0.7144	0.7770	0.7740

Tabulka 1: Korelace se všemi používanými příznaky (MEF vs. AEF)

Pro lepší představivost závislosti artikulačních příznaků a MFCC jsme však ještě spočítali korelaci, kdy jsme z příznaků odebrali energii a fundamentální frekvenci (viz tab. 2). Konkatenační ceny jsou tedy závislé jen na MFCC nebo na artikulačních příznacích.

Počet vět	**1**	**10**	**100**
Korelační koeficient	0.7988	0.7910	0.8002

Tabulka 2: Korelace pouze s MFCC (M vs. A)

Zde korelační koeficient ještě vzrostl, z čehož usuzujeme, že mezi MFCC a artikulačními příznaky existuje nějaký vztah blízký lineární závislosti.

4. Plánované experimenty

To, že korelační koeficient není roven jedné, by mohlo znamenat, že v artikulačních příznacích může být skryta nějaká informace, která v MFCC není a která by, pokud se nám podaří získat více dat, mohla vést ke zlepšení kvality. Jedním z našich dalších cílů je tedy nahrát hlas, který bude obsahovat více řečových dat. Získání artikulačních dat pomocí nahrávání však není příliš jednoduché, a proto bychom také do budoucna chtěli odhadovat artikulační parametry přímo z řečového signálu.

Poděkování

Příspěvek byl podpořen grantovým projektem SVK1-2017-021.

Literatura

Richmond, K., a King, S., 2016. Smooth Talking: Articulatory Join Cost for Unit Selection. *Proceedings, 41st Int. Conference on Acoustics, Speech and Signal Processing*.