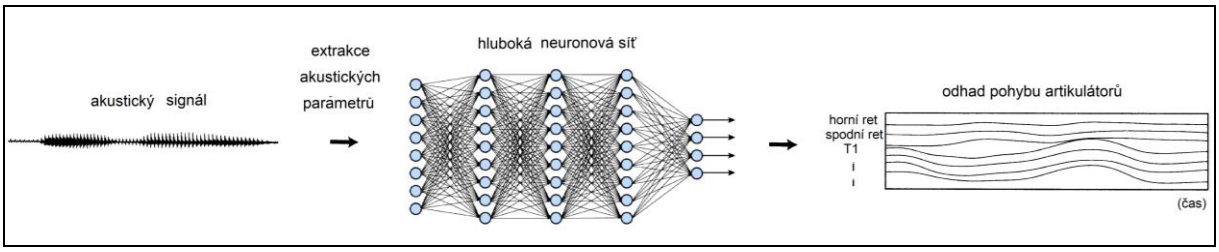
Akusticko-artikulační mapování

Martin Matura[1]

1 Úvod

Akusticko-artikulační mapování (inverze řeči) je proces, při kterém dochází k odhadování artikulačních trajektorií z akustického signálu. V dnešní době se pro tyto účely používají primárně hluboké neuronové sítě. Na jejich vstup přicházejí parametry akustického signálu (frekvence, energie, ...) a jejich výstupem jsou artikulační trajektorie (viz obr. 1).



Obrázek 1: Schéma predikce artikulačních trajektorií z akustického signálu

Inverze řeči se využívá především proto, že sběr artikulačních dat tradičním způsobem (ultrazvuk, magnetická rezonance, elektromagnetický artikulograf, ...) je náročný a to jak časově, tak technologicky. Artikulační trajektorie pak nacházejí využití například v syntéze či rozpoznávání řeči.

2 Predikční model

Jak bylo řečeno v úvodu, pro akusticko-artikulační mapování se využívají modely v podobě hlubokých neuronových sítí. V našich experimentech pracujeme s long-short-term-memory (LSTM) neuronovou sítí, což je typ rekurentní neuronové sítě, která se používá pro predikci sekvencí. Naše síť je tvořena vstupní vrstvou se 32 neurony, 3 skrytými vrstvami po 30 neuronech a výstupní vrstvou s 1 neuronem. Bylo provedeno 30 trénovacích epoch.

3 Příprava dat

Aby mapování mohlo proběhnout, je potřeba natrénovat predikční model. K tomu jsou potřeba trénovací data, která se skládají z akustické a artikulační části.

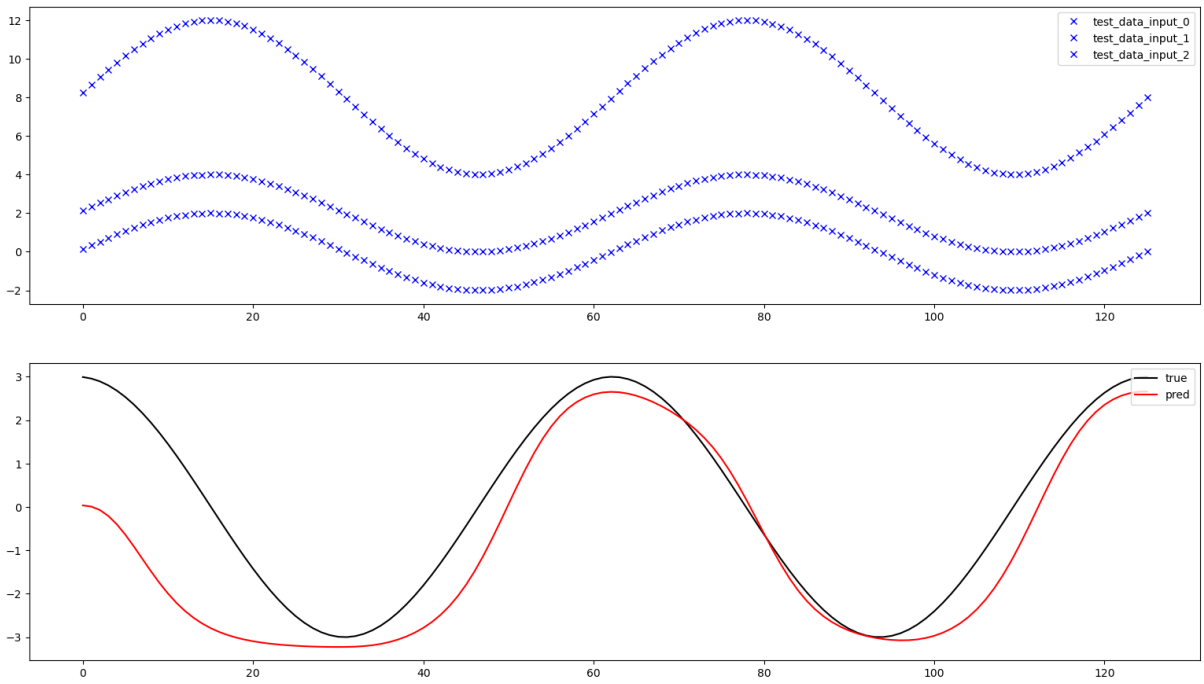
Pro sběr artikulačních dat jsme využili elektromagnetický artikulograf AG501, který umožňuje měřit artikulační trajektorie a synchronizovat je se zaznamenávaným řečovým signálem. Celý tento proces je však velmi pracný, protože senzory, které měří artikulační trajektorie, jsou připevněny na artikulátorech (rty, zuby, jazyk, ...) pomocí fyziologického lepidla a často se odlepují. Z toho důvodu se vždy pracuje s omezeným řečovým korpusem, který v našem případě obsahuje asi 800 krátkých vět.

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

Akustickou část dat (řečové signály) jsme popsali parametry - fundamentální frekvencí, energií a Melovskými kepstrálními koeficienty.

4 První experimenty

V oblasti akusticko-artikulačního mapování jsme provedli prvotní experimenty, které využívaly zjednodušených dat v podobě trigonometrických funkcí sinus a kosinus. Cílem bylo naučit hlubokou LSTM neuronovou síť predikovat ze tří vstupních funkcí sinus výstupní funkci kosinus. Jak naznačuje obrázek 2, náš naučený model byl poměrně dobře schopen predikovat (*pred*) tvar kosinové funkce (*true*).



Obrázek 2: Výsledky prvotního experimentu – nahoře vstup a dole výstup neuronové sítě (osy: *x – vzorky*, *y – amplituda*)

5 Závěr

V našem experimentu jsme použili jednoduchá data pro natrénování LSTM neuronové sítě a následně jsme predikovali požadovanou kosinovou trajektorii. V navazujících experimentech budeme pro trénování predikčního modelu využívat reálná akustická i artikulační data, která jsme získali měřením pomocí elektromagnetického artikulografu AG501.

Poděkování

Příspěvek byl podpořen grantovým projektem SGS-2019-027.

Literatura

Brownlee, J. (2020). A Gentle Introduction to LSTM Autoencoders. *Machine Learning Mastery, 27*. Available from: https://machinelearningmastery.com/lstm-autoencoders/