



Nahrávání největší databáze pohybových dat českého znakového jazyka

Pavel Jedlička¹

1 Úvod

Znakový jazyk (ZJ) je přirozený prostředek komunikace pro komunitu neslyšících. Ačkoliv je video považováno za přirozený způsob záznamu ZJ, pohyb se odehrává v 3D prostoru. Video je však pouze projekce prostorové scény do roviny. Technologie Motion Capture (česky lze přeložit jako Záznam pohybu, dále MoCap) umožňuje zaznamenávat pohybová data v podobě 3D trajektorií a tím umožňuje zaznamenat pohyb vhodněji, viz Dilsizian et al. (2016).

Cílem této práce je navázat na práci Jedlička et al. (2020), kde byl pomocí technologie MoCap pořízen tematicky omezený dataset ZJ. Ve zmiňované práci byl použit přístup pro záznam všech složek ZJ, kdy se nahrály všechny části těla najednou jedním zařízením MoCap s 18 kamerami. Celková doba záznamu je přibližně 42 minut (30 minut pro plynulé promluvy a 12 minut slovník izolovaných znaků), což tento dataset řadí jako největší MoCap dataset českého znakového jazyka (CSE) a objemem přibližně odpovídá největším MoCap datasetům napříč světovými jazyky. Představovaný postup má za cíl zjednodušit a urychlit proces nahrávání tak, aby bylo možné nahrát hodiny záznamu. Cílem je nahrát kolem 25 hodin záznamu a na rozdíl od předešlého datasetu se počítá s více různými řečníky.

2 Popis úlohy

Cílem práce je navrhnout postup pro vytvoření MoCap datasetu CSE. Práce vychází z předešlé práce Jedlička et al. (2020), kde byl použit systém s 18 kamerami a všechny složky projevu byly nahrány najednou. Tento přístup má velkou výhodu v konzistenci dat, ale je velice časově náročný na přípravu, kdy se scéna (rozmístění kamer) připravuje pro konkrétního řečníka a je velice citlivá na antropometrické parametry znakuujícího. Tato příprava je časově náročná a omezuje tak rychlost nahrávání. Při nahrávání nového datasetu se počítá s nahráváním projevů ZJ po částech. Zvlášť se budou nahrávat pohyby těla a paží a zvlášť tvary rukou. Důvodem je, že tyto dvě části vyžadují rozdílný přístup. Záběry na celé tělo vyžadují menší detail a zabírají větší prostor. Tvary rukou naopak vyžadují přesnější snímání a pohyby se odehrávají zpravidla v omezeném prostoru. Pro realizaci nahrávání bude použito 8 kamer, což představuje rovněž ekonomickou úsporu.

2.1 Definice datasetu

Dataset bude tvořen plynulými promluvami ve ZJ a slovníkovými položkami. V plynulých promluvách znaky na sebe přímo navazují a ovlivňují navzájem svou podobu. Slovníkové položky jsou naopak izolované znaky začínající a končící v klidové poloze rukou, v plynulých promluvách využívané pro ohrazení větných celků.

¹ student doktorského studijního programu Aplikované vědy a informatika, obor Kybernetika a řídicí technika, e-mail: jedlicka@students.zcu.cz

2.1.1 Témata

Jedním z plánovaných způsobů využití datasetu je datově řízená syntéza promluv v ZJ. Téma nahrávání je tedy vybráno tak, aby věty v něm použité měly pevnější strukturu a bylo možné nahrazením několika znaků v projevu změnit jeho význam. Zároveň se tímto přístupem můžeme vyhnout potřebě generovat komplexní gramatické jevy. Jedním z tématů je stejně jako v práci Jedlička et al. (2020) předpověď počasí. Dalším vybraným tématem je popis zvířat v ZOO. Popis se omezuje na faktické informace o místě výskytu, fyzických vlastnostech a základních informacích o životě. Takový popis by měl opět umožňovat při nahrazení některých výrazů generovat nové popisy. Jazyková kvalita, výběr znakujících a výběr témat je zajištěn spoluprací s lingvisty z Univerzity Palackého v Olomouci (UPOL), kteří během nahrávání projevů kontrolovali jazykovou čistotu projevu.

3 Závěr

Cílem práce je připravit podmínky pro realizaci nahrávání datasetu CSE. Byly provedeny experimenty s nahráváním pohybů těla a paží odděleně od tvarů ruky. Tímto způsobem lze efektivně pořídit větší množství dat než nahráváním najednou. Nevýhodou tohoto přístupu je nutnost spojit odděleně nahraná data. Informace v datech pohybů těla obsahují informaci o natočení zápěstí (orientace ruky). Data tvaru ruky obsahují informace o pohybech prstů. Data lze tedy snadno spojit z hlediska prostoru. Problém ovšem nastává u časové synchronizace, vyřešení této otázky je zásadní pro přirozenost a srozumitelnost výsledné syntézy.

Poděkování

Příspěvek byl podpořen grantovým projektem číslo SGS-2019-027.

Literatura

Dilsizian, M., Tang, Z., Metaxas, D., Huenerfauth, M., Neidle, C. *The Importance of 3D Motion Trajectories for Computer-based Sign Recognition* 7th Workshop on the Representation and Processing of Sign Languages: Corpus Mining, The 10th International Conference on Language Resources and Evaluation (LREC 2016).

Jedlička, Pavel and Krňoul, Zdeněk and Kanis, Jakub and Železný, Miloš, *Sign Language Motion Capture Dataset for Data-driven Synthesis*, Proceedings of the LREC2020 9th Workshop on the Representation and Processing of Sign Languages: Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives, 12th International Conference on Language Resources and Evaluation (LREC 2020)