

Západočeská univerzita v Plzni  
Fakulta aplikovaných věd  
Katedra kybernetiky

# BAKALÁŘSKÁ PRÁCE

Plzeň, 2020

Tomáš Železný

## PROHLÁŠENÍ

Předkládám tímto k posouzení a obhajobě bakalářskou práci zpracovanou na závěr studia na Fakultě aplikovaných věd Západočeské univerzity v Plzni.

Prohlašuji, že jsem bakalářskou práci vypracoval samostatně a výhradně s použitím odborné literatury a pramenů, jejichž úplný seznam je její součástí.

v Plzni dne

.....

### **Poděkování**

Rád bych poděkoval svému vedoucímu bakalářské práce Ing. Pavlu Jedličkovi za cenné rady, připomínky a konzultace. Dále bych chtěl poděkovat Ing. Zbyňku Tychtlovi, Ph.D. za cenné konzultace v oblasti frekvenční analýzy.

## **Anotace**

Znakový jazyk je důležitým komunikačním prostředkem Neslyšících. Protože ne vždy je při překladu do znakového jazyka jednoduché použít tlumočení, je snahou vytvořit jeho automatickou syntézu. Protože je pro její vznik potřeba velké množství dat, jejichž pořízení je velmi časově náročné, je nutné tato data vytvořit procesem augmentace. Pro tato umělá data je potřeba posoudit jejich přirozenost. Rychlého objektivního hodnocení lze dosáhnout pomocí automatických metod určující podobnost a jiné parametry, které by při určení přirozenosti mohly rozhodovat. Tato práce pojednává o výběru metod a měřitelných veličin, které by k tvorbě augmentovaných dat mohly přispět.

## **Klíčová slova**

Záznam pohybu, znakový jazyk, analýza dat, frekvenční analýza, augmentace dat

## **Abstract**

Sign language is an important mean of communication of the Deaf. Because it is not always easy to use interpreting when translating into sign language, there is an effort to create an automatic synthesis. Large amount of data is required for its creation. The acquisition of such data is very time consuming so it is necessary to create additional data through the augmentation process. For such artificial data, is necessary to evaluate its naturalness. Rapid objective evaluation can be achieved by automatic methods determining similarity and other parameters that could decide in determining naturalness. This work deals with the selection of methods and measurable quantities that could contribute to the creation of augmented data.

## **Key words**

Motion capture, sign language, data analysis, frequency analysis, data augmentation

# Obsah

<b>1</b>	<b>Úvod</b>	<b>1</b>
<b>2</b>	<b>Znakový jazyk</b>	<b>2</b>
2.1	Historie . . . . .	2
2.2	Popis jazyka . . . . .	3
2.2.1	Prostor . . . . .	3
2.2.2	Specifické znaky . . . . .	3
2.3	Notace . . . . .	5
2.3.1	Stokoeho notační systém . . . . .	5
2.3.2	SignWriting . . . . .	6
2.3.3	HamNoSys . . . . .	6
<b>3</b>	<b>Záznam pohybu</b>	<b>8</b>
3.1	Historie . . . . .	8
3.2	Princip . . . . .	10
3.2.1	Mechanické systémy . . . . .	11
3.2.2	Elektromagnetické systémy . . . . .	11
3.2.3	Inerciální systémy . . . . .	11
3.2.4	Optické systémy . . . . .	12
3.3	Postprocessing . . . . .	13
<b>4</b>	<b>Data</b>	<b>16</b>
4.1	Motivace . . . . .	17
4.2	Pořízení dat . . . . .	17
4.3	Struktura dat . . . . .	17
4.4	Práce s daty . . . . .	19
<b>5</b>	<b>Analýza</b>	<b>22</b>
5.1	Pozice . . . . .	22
5.2	Rychlost a zrychlení . . . . .	22
5.3	Dynamika pohybu . . . . .	23
5.4	Dominance ruky . . . . .	23
5.5	Určení jednorukosti-dvourukosti . . . . .	24
5.6	Vyhodnocení podobnosti . . . . .	25
5.7	Symetrie znaků . . . . .	29
5.8	Frekvenční analýza . . . . .	32
<b>6</b>	<b>Závěr</b>	<b>38</b>

# 1 Úvod

Pojem znakový jazyk (ZJ) označuje vizuálně motorický komunikační prostředek používaný na celém světě. ZJ se tedy vyjadřuje pomocí pohybu, pozic a tvaru rukou v prostoru, pohybech těla a hlavy včetně mimiky. Je používán lidmi, kteří mají důvod nepoužívat verbální komunikaci. Zejména je tento jazyk upřednostňován lidmi, kteří mají problém s vnímáním zvuku a těmi kteří s nimi komunikují. Tito lidé bývají rozdělováni podle druhu vady sluchu na nedoslýchavé - ti, kteří mají částečnou ztrátu sluchu, ale za pomoci správné techniky či prostředí mohou rozumět mluvené řeči, neslyšící - ti, kteří nemohou rozumět mluvené řeči za žádných okolností a na ohluchlé - ti, kteří se stali neslyšícími až poté, co se naučili přirozeným způsobem mluvený jazyk. Těm, kteří ztratili sluch před tím, než se naučili přirozeným způsobem mluvený jazyk nebo se již narodili neslyšící, se říká prelingválně neslyšící. Komunita lidí, kteří se cítí být členy menšiny preferující komunikace ve ZJ, se nazývá Neslyšící [1]. ZJk nemá psanou formu. s mluveným jazykem v dané oblasti, tedy např. český ZJ s češtinou, nemá mnoho společného. Pro lidi, kteří mají ZJ jako mateřský, je tedy přirozenější komunikace ve ZJ než ve psané podobě mluveného jazyka. Ta je pro ně jako cizí jazyk.

Protože ne vždy je při tlumočení předem známa promluva a proces simultánního tlumočení není triviální (viz kapitola 4.1), je snaha vyvinout automatickou syntézu ZJ, která Neslyšícím umožní přístup k většímu množství informací v přirozené formě. V současné době je nejčastější způsob syntézy ZJ založen na pravidlech. Tento typ syntézy není náročný na množství potřebných dat, není však dostatečně přirozený. Pro zajištění co nejpřirozenější syntézy je snaha ji tvořit za pomoci strojového učení. K trénování budou použita data ze zařízení záznamu pohybu, které zaznamenává pohyb přímo (viz kapitola 3). Tvorba takovéto syntézy vyžaduje velké množství vstupních dat, jejichž pořízení je však časově i organizačně velmi náročné, v praxi jich je tak velký nedostatek. Je proto snahou vytvořit umělá data na základě reálných tak, aby byly co nejpřirozenější a bylo by je možné použít jako vstupní data pro tvorbu syntézy. Tento proces se nazývá augmentace. Vytvořit augmentaci takovou, aby splňovala předpoklady lidské přirozenosti, není jednoduché. Metody augmentace jsou závislé na typu dat a jejich dalším zpracování. Data je proto třeba důkladně analyzovat, aby výsledná augmentace odpovídala realistickým datům.

Cílem této práce je proto provést analýzu dat pro různé druhy následné augmentace. Jednou z nejběžnějších je augmentace zašuměním původních dat. Protože však není jasné, jaký typ šumu použít, aby si data zachovaly přirozenost, bude pro použitá data provedena frekvenční analýza. Jejím cílem bude oddělení frekvencí šumu a frekvencí nesoucí informaci. Změnou dat na vhodných frekvencích tak nepoškodí nesenou informaci. Dalším způsobem augmentace je mutace vybraných částí těla pro jednotlivé znaky navzájem. Pro zhodnocení těchto uměle vytvořených znaků je třeba metricky, která hodnotí přirozenost. Jediný, kdo v principu může rozhodnout o přirozenosti lidských pohybů je člověk. Jedná se však o subjektivní hodnocení a pro jeho objektivnění by bylo třeba příliš velkého počtu hodnocení, které by bylo velmi časově náročné. Pro potřeby strojového učení je třeba objektivního hodnocení. Automaticky sice není možné určit přirozenost pohybů, ale je možné určit jejich podobnost a další parametry, na jejichž základě by bylo možné přirozenost hodnotit. Základními veličinami, které lze pro použitá data zkoumat, jsou rychlost a zrychlení jednotlivých částí těla, na jejichž základě lze odvodit další veličiny charakterizující pohyb a na nichž je možné hodnotit jeho kvalitu.

## 2 Znakový jazyk

Pod pojmem znakový jazyk je označována řada vizuálně motorických jazyků, které používají lidé po celém světě. Tyto jazyky jsou plnohodnotnými jazyky Neslyšících. Stejně jako mluvené jazyky se znakový jazyk stále vyvíjí a je v různých regionech odlišný. Není však žádným způsobem na mluveném jazyku závislý, Český znakový jazyk se tedy od češtiny liší jak slovníkem tak gramatikou.

V České republice nebyl český znakový jazyk dlouho uznávaný. Neslyšící byli nuceni používat znakovou češtinu. Jedná se o převedenou mluvenou češtinu do znaků. Byla uměle vytvořena slyšícími s cílem snadného a přesného tlumočení z češtiny. Přejímá gramatiku češtiny a slovník z části ze znakového jazyka, pro slova, která nemají znak je využívána prstová abeceda. Tento způsob je však velice zdlouhavý, náročný a především pro Neslyšící nepřirozený. Český znakový jazyk začali podle [2] jazykovědci studovat až v 90. letech 20. století. v roce 1998 byl přijat zákon o znakovém jazyce (Zákon č. 155/1998 Sb. Zákon o komunikačních systémech neslyšících a hluchoslepých osob), který Neslyšícím zaručuje právo na používání znakového jazyka a vzdělávání se v něm.

### 2.1 Historie

Nejstarší historii znakového jazyka popisuje [3]. Podle něj se o prstové abecedě poprvé zmínil anglosaský mnich Beda Venerabilis, který žil v letech 672-735. v 10. století, když se v kláštorech začali mniši zavazovat trvalým mlčením, používali ke komunikaci mezi sebou právě prstovou abecedu. Později začali používat jednoduchou znakovou řeč. v 16. století znakovou řeč poprvé začal vyučovat Neslyšícím španělský mnich Pedro Ponce de León, který využíval své zkušenosti se znakovou řečí mezi mnichy v klášteře. v roce 1620 španělský šlechtic Juan Martin Pablo Bonet napsal knihu o vyučování Neslyšících. Jako první popsal jednoruční prstovou abecedu. Anglický lékař John Bulwer pak k dosavadním prostředkům připojil metodu odezírání ze rtů.

Na základě všech znalostí o výuce znakového jazyka vypracoval francouzský kněz Abbé Charles Michel de l'Epée metodu vyučování Neslyšících. v roce 1770 založil školu pro Neslyšící. Tato škola měla velký úspěch a žáků přibývalo. Zabýval se také výukou učitelů, kteří poté zakládali své školy pro Neslyšící. v roce 1776 vydal první knihu o znacích a prstové abecedě. Jeho způsob výuky se říká francouzská metoda. Dodnes je Abbé l'Epée pokládán za Otce Neslyšících. Jeho žák Clerc přenesl tento způsob použití znakového jazyka do USA, kde se stále používá.

Francouz J.Rodriguez Pereira učil svoji neslyšící sestru artikulaci a odezírání řeči s použitím znaků. Začal také rozlišovat různé stupně hluchoty. Své zkušenosti předložil francouzské akademii. z německých zemí se dostalo kritiky francouzské metodě. Byly zde odmítány znaky a byla spíše propagována orální metoda. Lékař Johann Konrad Amman převzal metody pro artikulaci od Španělů a odezírání od Johna Bulwera. Ve výuce nejprve používal artikulaci pak teprve písmo. Němec Ferdinand Arnoldi navázal na Ammanovu metodu a připojil k ní psaní. Vyučoval němčinu, počty a náboženství.

Roku 1880 byl během Milánského kongresu zamítnut znakový jazyk jako vyučující jazyk Neslyšících. Orální metoda se stala nadřazenou znakovému jazyku. Neslyšící se roku 1900 na protest sešli na konferenci v Paříži a odmítli vpustit dovnitř oralisty. Přesto byli poraženi a další protesty nepořádali. v 60. letech 20. století prokázal americký jazykovědec William Stokoe, že znakový jazyk má rysy spojované s lidskými jazyky. Ukázal, že tvary, pozice a pohyby znakového jazyka jsou uspořádány do lin-

gvistických struktur s podobnými funkcemi, jako má mluvený jazyk. Od té doby se ve světě začala rozvíjet lingvistika znakových jazyků. Hluchota přestala být považována za poruchu, která je třeba napravit odezíráním ze rtů.

## 2.2 Popis jazyka

Znakový jazyk má dvě hlavní složky. Manuální složkou jsou pohyby, tvary a pozice rukou. Nemanuální složkou jsou pohyby a pozice hlavy, horní části trupu a mimika. Tyto složky jsou nedílnou součástí znakového jazyka. Mohou být produkovány a vnímány současně, například mimika bývá používána pro vyjádření emocí. Dva jednoruční znaky pak mohou být znakovány vedle sebe ve složeninách.

Na rozdíl od gest lze znaky analyzovat, rozložit na jednotlivé komponenty jako lze dělit slova. Tyto komponenty však nejsou poskládány za sebe, jako je tomu u mluveného jazyka, ale mohou se vrstvit přes sebe. Znakové jazyky tedy nemají psanou podobu. Pro zaznamenání znaků se používají notační systémy [2.3].

Znaky se podle [4] dají rozdělit na znaky artikulované jednou rukou (dále jednoruční), znaky artikulované dvěma rukama (dále dvouruční) a znaky složené. Jednoruční znaky jsou artikulované pomocí dominantní ruky, zatímco druhá ruka není využívána. Dominantní ruka je ta, kterou mluvčí upřednostňuje. Dvouruční znaky mohou mít buď obě ruce aktivní, nebo jednu aktivní a druhou pasivní, například aktivní něco ukazuje na pasivní ruce (příklad je na Obrázku 1). Složené znaky jsou kombinace více znaků.

### 2.2.1 Prostor

Ve znakovém jazyce hraje velkou roli prostor. Dle [5] lze prostor dělit na typografický a prostor syntaktický. Topografický prostor odpovídá reálnému prostoru v čase promluvy, například rozmístění předmětů v místnosti. Prostor syntaktický se liší od reálného, mluvčí si volí uspořádání objektů v prostoru. To je například typické pro vyprávění. v obou typech prostoru dochází ke změně držení těla a výrazu obličeje. Mluvčí situaci produkuje tak, jakoby se vtělil do předváděného objektu. Dochází k tomu například, když vypráví rozhovor dvou osob. Mění postoj těla a výraz v obličeji podle předváděných osob a přestane vázat oční kontakt s adresátem.

Pomocí prostoru může být také vyjadřován čas. Existují časové osy, na nichž může být vymezen časový rámeček, v němž se děj odehrává. Základní osa se nachází nad ramenem dominantní ruky. Na této ose se mluvčí vyjadřuje, kdy se děj koná. Co se již odehrálo leží za zády mluvčího (např. znak "včera") a co se teprve odehraje leží před mluvčím (např. znak "zítra"). Na Obrázku 2 jsou vyobrazeny časové linie používané v britském znakovém jazyce pro dominantní pravou ruku. Zmíněná osa je symbolizovaná linií A. Přesně vymezené kalendářní doby se artikulují dopředu od těla. Na obrázku je tato osa znázorněna linií B. Zleva doprava se před tělem označuje časový sled či trvání události. Osa je znázorněna linií C. Odspoda nahoru je značena doba růstu či dospívání. Je znázorněna linií D.

### 2.2.2 Specifické znaky

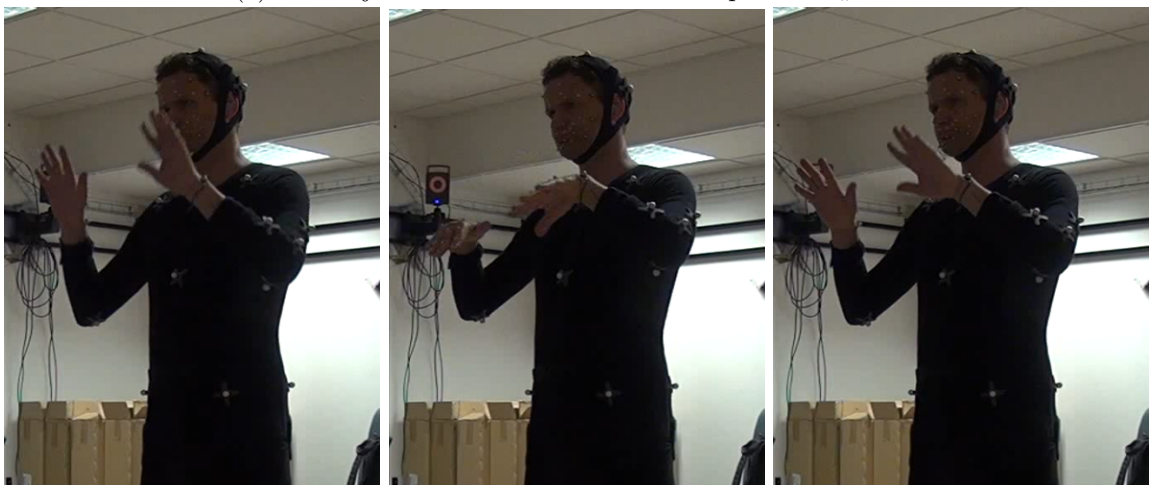
Ve slovní zásobě znakových jazyků se objevují znaky, které nelze jednoslovně přeložit do českého jazyka. [6] tyto znaky nazývá jako specifické znaky. Jedná se o znaky s čistě

---

<sup>1</sup>Zdroj: <http://ruce.cz/clanky/546-poznavame-cesky-znakovy-jazyk-iv>

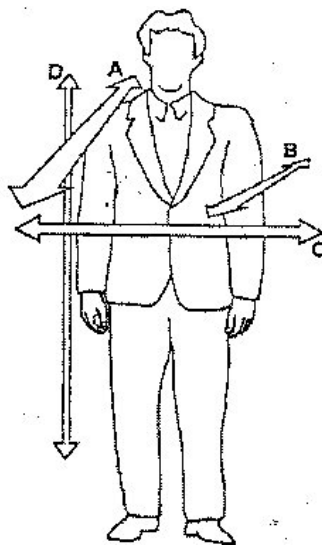


(a) Znak s jednou rukou aktivní a druhou pasivní - „metr“



(b) Znak s oběma rukama aktivními - „pršet“

Obrázek 1: Příklad znaku s jednou rukou pasivní a druhou aktivní a znaku s oběma rukama aktivními



Obrázek 2: Časové linie v britském znakovém jazyce<sup>1</sup>



nemanuální složkou, které se kombinují spolu s manuálními znaky. Nejčastěji vyjadřují nějaké emoce, postoj mluvčího k dané věci či jeho názor. Dle [7] jsou specifické znaky součástí gramatiky znakového jazyka a patří do mateřského jazyka Neslyšících. Ti je používají velmi často, někdy se objevuje více specifických znaků v jedné promluvě.

Specifické znaky jsou sdělovány najednou s manuálními znaky a lze je obtížně přeložit do mluveného jazyka. Většina Neslyšících při komunikaci se slyšícím člověkem vynechává specifické znaky kvůli jejich těžké srozumitelnosti. Proto není snadné tyto znaky zachytit. Jedná se však o přirozený jev komunikace rodilého mluvčího libovolného jazyka s cizincem nazývaným *foreigner talk* [8].

## 2.3 Notace

Jak již bylo zmíněno dříve, znakový jazyk nemá psanou formu. Aby ho bylo možné nějakým způsobem popsat, hledali lingvisté způsob jakým ho zaznamenat. v průběhu času bylo vymyšleno více tzv. notačních zápisů, některé z nich jsou popsány níže. Většina z nich se zaměřuje na nějakou konkrétní část znakového jazyka, avšak opomíná jiné části. Nejjednodušším zápisem bývají tzv. glosy. Jedná se o text, kde každý znak zastupuje jeho přepis ve psaném jazyce. Tyto glosy mohou vyvolávat dojem, že má znakový jazyk psanou formu. Neumožňují však zápis flexe znaků (např. skloňování, časování), nebo nemanuální složky. Nejlepším způsobem záznamu znakového jazyka je video, protože obsahuje veškeré informace o promluvě.

### 2.3.1 Stokoeho notační systém

William Stokoe byl americký jazykovědec, který prokázal, že složky znakového jazyka jsou uspořádány do lingvistických struktur srovnatelných s mluveným jazykem. Ve své studii *Sign Language Structure* [4] rozdělil znak na tzv. cherémy (složenina řeckého kořene *cheir* - ruka a slova foném). Znak se skládá ze tří cherémů:

1. TAB - Pozice znaku v prostoru (tabula)
2. DEZ - Tvar ruky/rukou a orientace (designator)
3. SIG - Pohyb ruky/rukou (signator)

Tyto cherémy jsou produkovány současně, nelze je v čase oddělit. Jsou zapisovány symboly založenými na latinské abecedě lineárně za sebou jako psaný text. Na rozdíl od cherémů není nutné, aby čtenář notace musel znát význam znaku aby ho dokázal zreprodukovat. Na Obrázku 3 znázorněn příklad zápisu Stokoeho notačního systému.



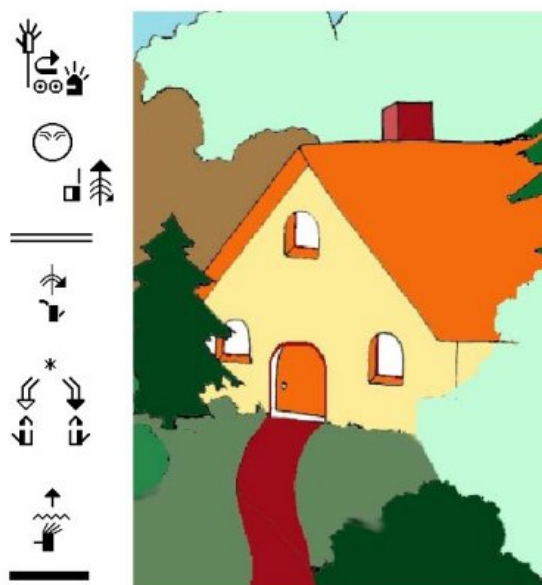
Obrázek 3: Stokoeho notace pro znak amerického znakového jazyka „don't know“<sup>2</sup>

<sup>2</sup>Zdroj: [https://www.researchgate.net/figure/Stokoe-notation-for-the-sign-dont-know-in-ASL\\_fig1\\_235760493](https://www.researchgate.net/figure/Stokoe-notation-for-the-sign-dont-know-in-ASL_fig1_235760493)

### 2.3.2 SignWriting

V roce 1973 Američanka Valerie Sutton vytvořila systém pro zápis baletního tance DanceWriting. v té době se se znakovým jazykem nikdy nesetkala. v roce 1974 na žádost dánského lingvisty Larse von der Lietha modifikací DanceWritingu vytvořila systém zápisu znakového jazyka SignWriting [9].

Zápis SignWritingu je proti Stokoeho notaci více intuitivní, liší se symboly i tím, že zápis není lineární. v notaci se vyskytuje velké množství symbolů v několika kategoriích. Mezi tyto kategorie patří například symboly reprezentující tvary rukou, pohyby, kontakty, symboly reprezentující hlavu či držení těla. Tyto symboly jsou zapisovány pod sebe do sloupců. Zápis nemusí obsahovat všechny kategorie. Symbolickými náčrtky lze zapsat jakýkoliv pohyb. Díky své čitelnosti pro člověka je hodně dalších typů notací založeno právě na notaci SignWriting, pro strojové učení je však špatně použitelný. Na Obrázku 4 je znázorněn příklad zápisu znakové řeči v SignWritingu.



Daleko daleko uprostřed lesa,  
na kopci stojí chaloupka.

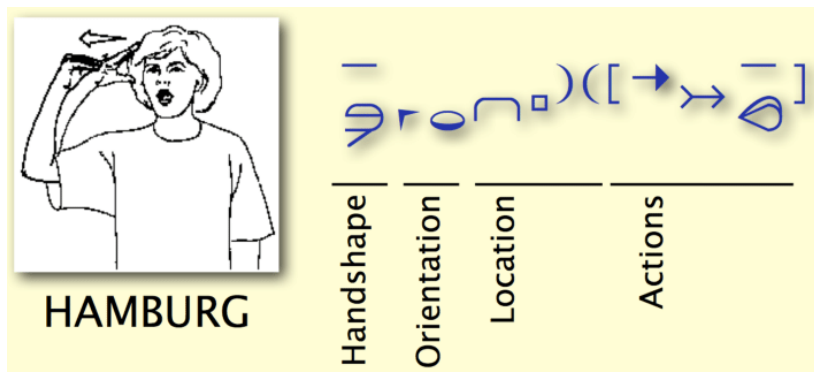
Obrázek 4: SignWriting notace pro část pohádky O Mášence a třech medvědech<sup>3</sup>

### 2.3.3 HamNoSys

The Hamburg Notation System for Sign Languages (HamNoSys)[10] je notační systém vytvořený Němcem Thomasem Hankem v roce 1985. Jedná se o zobecněnější obdobu Stokoeho notačního systému. Symboly jsou navrženy tak, aby byly intuitivní. Dělí se do 4 skupin: tvar ruky, orientace, lokace a pohyb. Jazykovědci na univerzitě v Hamburku stále pracují na zdokonalení tohoto notačního systému. Od doby vzniku prošel systém několika úpravami, například byl v průběhu doplněn o skupinu se symboly pro zaznamenání nemanuální složky.

<sup>3</sup>Zdroj: [http://galerie.ruce.cz/displayimage.php?album=5&pid=58#top\\_display\\_media](http://galerie.ruce.cz/displayimage.php?album=5&pid=58#top_display_media)

Díky své dobře zpracovatelné struktuře je HamNoSys jeden z nejpoužívanějších notačních systémů pro strojové zpracování znakového jazyka. Na druhou stranu je však oproti jiným notačním systémům špatně čitelný pro člověka. Přestože je v nejnovější verzi doplněn o symboly s nemanuální složkou, stále neobsahuje všechny možnosti této komponenty. Na Obrázku 5 je příklad zápisu notace HamNoSys.



Obrázek 5: Notace HamNoSys pro znak „Hamburg“<sup>4</sup>

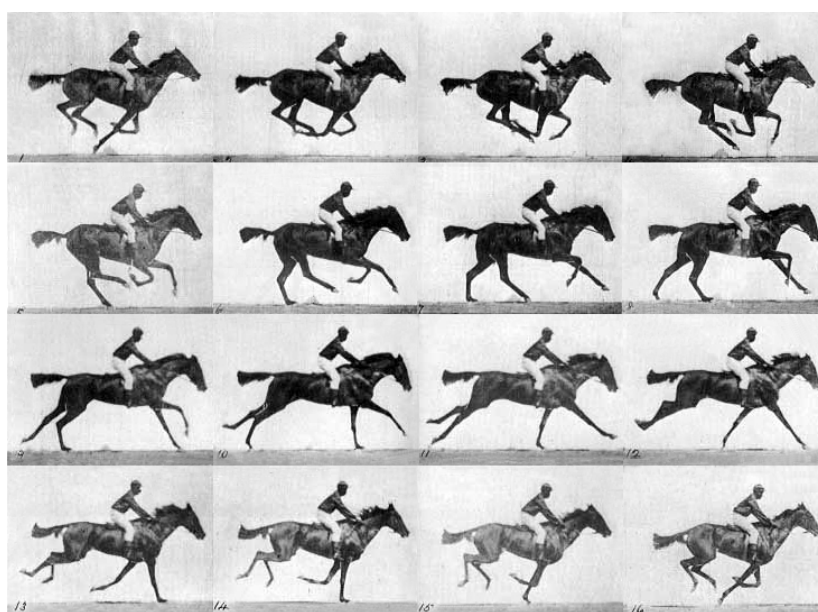
<sup>4</sup>Zdroj: [https://www.researchgate.net/figure/An-example-of-the-HAMBURG-sign-transcribed-in-HamNoSys-source-Hanke-2010\\_fig26\\_331938999](https://www.researchgate.net/figure/An-example-of-the-HAMBURG-sign-transcribed-in-HamNoSys-source-Hanke-2010_fig26_331938999)

### 3 Záznam pohybu

Pojmem záznam pohybu (anglicky motion capture nebo mezinárodně používanou zkratkou MoCap) je obecně nazýván proces nahrávání pohybu objektů. Nejčastěji bývá spojován se zařízením, které toto nahrávání umožňuje. Využití mívá například ve filmovém a herním průmyslu, ve zdravotnictví nebo jiných vědních oborech.

#### 3.1 Historie

Počátky záznamu pohybu se datují již do 19. století, kdy se kalifornský guvernér Leland Stanford v roce 1873 vsadil o 25000 \$, že se kůň při cvalu v určitou dobu nedotýká země žádnou nohou [11]. Anglo-americký vynálezce Eadweard Muybridge, který měl tento výrok potvrdit, pravidelně rozmístil 24 fotografických přístrojů, ke kterým vedl provázek tak, že je běžící kůň přetrhl a spustil závěrky jednotlivých přístrojů. Na Obrázku 6 jsou vidět výsledky tohoto experimentu, které potvrdily Stanfordův názor.



Obrázek 6: Muybridgeovo snímky koňského cvalu<sup>5</sup>

V roce 1882 francouzský fyziolog Étienne-Jules Marey nahradil Muybridgeovo vybavení svojí jednoduchou kamerou, která dokázala zachytit na jediný obrázek sérii snímků s rychlostí až 12 snímků za sekundu. Byla vybavena teleobjektivem, který sloužil ke snímání objektů na delší vzdálenosti. Často bývala vybavena ramenní podpěrkou a díky tomu připomínala pušku, podle které se jí říká fotografická puška (viz Obrázek 7).

Dalším vývojovým stupněm bylo vynalezení rotoskopu polsko-americkým filmařem Maxem Fleischerem v roce 1915. Rotoskopie je technika, kdy animátor používá nahrané snímky jako předlohu pro animaci. Je to zdoluhavý proces, protože je třeba překreslovat každý snímek. Animace je však díky reálné předloze souvislejší a realističtější. První Fleischerova animace byl pohyb klauna Koko podle jeho bratra Davida v klaunském

<sup>5</sup>Zdroj: [https://commons.wikimedia.org/wiki/File:Muybridge\\_race\\_horse\\_gallop.jpg](https://commons.wikimedia.org/wiki/File:Muybridge_race_horse_gallop.jpg)

<sup>6</sup>Zdroj: [https://upload.wikimedia.org/wikipedia/commons/7/7f/Fusil\\_de\\_Marey\\_p1040353.jpg](https://upload.wikimedia.org/wikipedia/commons/7/7f/Fusil_de_Marey_p1040353.jpg)

<sup>7</sup>Zdroj: [https://upload.wikimedia.org/wikipedia/commons/e/e0/Marey\\_-\\_birds.jpg](https://upload.wikimedia.org/wikipedia/commons/e/e0/Marey_-_birds.jpg)



Obrázek 7: Mareyova fotografická puška (a)<sup>6</sup> a snímek pelíkána na ni zachycený (b)<sup>7</sup>

kostýmu (Obrázek 8). Prvním celovečerním filmem animovaným pomocí rotoskopie byl v roce 1937 *Snow White and the Seven Dwarfs* od Walta Disneyho. v dnešní době je rotoskopie stále využívána jako referenční materiál pro animátory.



Obrázek 8: Fleischerův klaun Koko<sup>8</sup>

S příchodem digitálních technologií a dostatečného výpočetního výkonu začaly vznikat 3D animace. Jednou z prvních animací je krátký spot s robotem Brilliance z roku 1985 [12]. Jeho autor Robert Abel vycházel pouze z černých bodů umístěných na těle herce nasnímaných fotoaparáty. Mezníkem bylo použití počítače pro účely zpracování takovýchto fotografií. Prvním charakterem celovečerního filmu vytvořeným uměle pomocí MoCapu byl *Jar Jar Bings* hraný Ahmedem Bestem ze *Star Wars: Episode I - The Phantom Menace* [13]. První celovečerní film, který využil záznam pohybu v reálném čase byl *The Lord of the Rings: The Two Towers* [14], kde herec Andy Serkis ztvárnil postavu Gollum (viz Obrázek 9). Serkisovy pohyby byly v reálném čase přenášeny do počítače, kde se přenášely na generovanou postavu. v roce 2009 byl použit MoCap ve formě helmy pro záznam pohybu obličeje ve filmu *Avatar* [15] (viz Obrázek 10).

<sup>8</sup>Zdroj: [https://upload.wikimedia.org/wikipedia/commons/thumb/4/44/Tantalizing\\_Fly\\_1919.AVI.025s40to85q6a5mono.ogv/260px--Tantalizing\\_Fly\\_1919.AVI.025s40to85q6a5mono.ogv.jpg](https://upload.wikimedia.org/wikipedia/commons/thumb/4/44/Tantalizing_Fly_1919.AVI.025s40to85q6a5mono.ogv/260px--Tantalizing_Fly_1919.AVI.025s40to85q6a5mono.ogv.jpg)

<sup>9</sup>Zdroj: <http://blog.castac.org/2019/03/out-of-body-workspaces-andy-serkis-and-motion-capture-technologies/>

<sup>10</sup>Zdroj: <https://srushtivfx.com/how-digital-characters-replicate-human-action-and-emotions-motion-capture/>



Obrázek 9: Gollum<sup>9</sup>



Obrázek 10: Avatar<sup>10</sup>

### 3.2 Princip

Moderní MoCap zařízení jsou obecně využívána pro snímání pohybu libovolného objektu. Ty se dají rozdělit na dvě skupiny; tuhý předmět (rigid) je například meč nebo tělo dronu a ohebný (soft) je například člověk. MoCap systémy lze rozdělit podle způsobu, kterým snímají pohyb (viz níže). Obecně v průběhu času zaznamenávají hodnoty daných veličin, ze kterých jsou následně vypočteny trajektorie polohy či rotace a z nich je poté rekonstruován digitální model objektu. Před zahájením záznamu je třeba vždy zkalibrovat příslušné zařízení. Offline systémy během záznamu ukládají pouze hodnoty těchto veličin a model rekonstruují až po ukončení záznamu. Online zařízení dokážou produkovat data v reálném čase, za cenu kvality a rychlosti přenosu.

#### Vlastnosti

Při volbě druhu zařízení pro MoCap, který pro danou úlohu použít, hraje roli několik vlastností:

- Přesnost udává maximální chybu, která může vzniknout při snímání polohy či rotace.
- Rozptyl udává odchylku, se kterou hodnoty kolísají kolem skutečné hodnoty. Čím větší je rozptyl, tím méně plynulé pohyby získáme. Data tzv. šumí.
- Latence udává dobu zpoždění, s jakou systém zareaguje na pohyb objektu.
- Snímkovací frekvence udává kolikrát zvládne systém změřit polohu či rotaci objektu za jednotku času.
- Přirozenost pohybu objektu. Snižuje se v závislosti na zařízeních nutných k nahrávání, která omezují v pohybu, např. exoskelet či přírodní kabel. Dále také závisí na počtu použitých senzorů či značek. Čím je jich více, tím více informací o pohybu získáme a je tak přirozenější.
- Znalost absolutní pozice. v případě že ji neznáme, je obtížné nahrávat pohyb, kdy se objekt nedotýká podlahy, tedy např. skok nebo běh.
- Cena



### 3.2.1 Mechanické systémy

Pohyb objektu je snímán pomocí mechanického senzoru. v případě objektů, kde není možné měřit ohyb přímo, tedy např. člověk, je třeba na daný objekt připevnit exoskelet. Jedná se o systém kovových či plastových tyčí, které jsou spojeny klouby, ve kterých jsou umístěny senzory. Pomocí těchto senzorů je pak snímán pohyb objektu. Protože tento exoskelet nikdy nepasuje úplně přesně na snímáný objekt, není pohyb a jeho rozsah přirozený. Mechanické systémy se někdy z těchto důvodů používají pouze pro části těla, např. MoCap rukavice pro záznam pohybu prstů. Data neobsahují informace o orientaci nebo kontaktu s jinými objekty. Na druhou stranu je tento typ poměrně přesný a levný. Příklad mechanického systému využívající exoskeleton je na Obrázku 11.



Obrázek 11: Mechanický systém<sup>11</sup>

### 3.2.2 Elektromagnetické systémy

Na snímáný objekt je umístěna řada magnetických přijímačů (viz Obrázek 12). Ty vypočítávají pozici a orientaci díky relativnímu magnetickému toku cívek umístěných na objektu a statického generátoru magnetického pole. Data obsahují informaci o absolutní pozici a orientaci. Na druhou stranu nejsou data tolik přesná a bývají zašuměná. Překrytí senzorů částí těla nemá na nahrávání takový vliv, jako má u optických systémů. Používají se tedy například pro snímání pohybu jazyka tzv. artikulografem. Elektromagnetické systémy se často používají v kombinaci s jinými metodami.

### 3.2.3 Inerciální systémy

Tato technologie je založena na kombinaci gyroskopů, akcelerometrů a magnetometrů rozmístěných na snímáném objektu. Data ze všech těchto senzorů jsou shromažďována do počítače, kde jsou zpracovávána [16]. Nevýhodami jsou malá flexibilita,

<sup>11</sup>Zdroj: <https://baswaramursyid.wordpress.com/2015/01/20/motion-capture-ii/>

<sup>12</sup>Zdroj: [https://www.researchgate.net/figure/3-A-performer-wearing-a-suit-for-magnetic-motion-capture-AMM-2010\\_fig4\\_255990108](https://www.researchgate.net/figure/3-A-performer-wearing-a-suit-for-magnetic-motion-capture-AMM-2010_fig4_255990108)



Obrázek 12: Elektromagnetický systém<sup>12</sup>

nížká přesnost a absence informace o absolutní pozici. Pro získání informace o absolutní poloze, se často používají kombinace s jiným typem přístupu, který tuto informaci poskytuje. Inerciální druh systému se používá tam, kde nejsou vysoké nároky na přesnost.

### 3.2.4 Optické systémy

Optické systémy záznamu pohybu používají ke zjištění polohy objektu kamery. Tyto systémy se dají rozdělit:

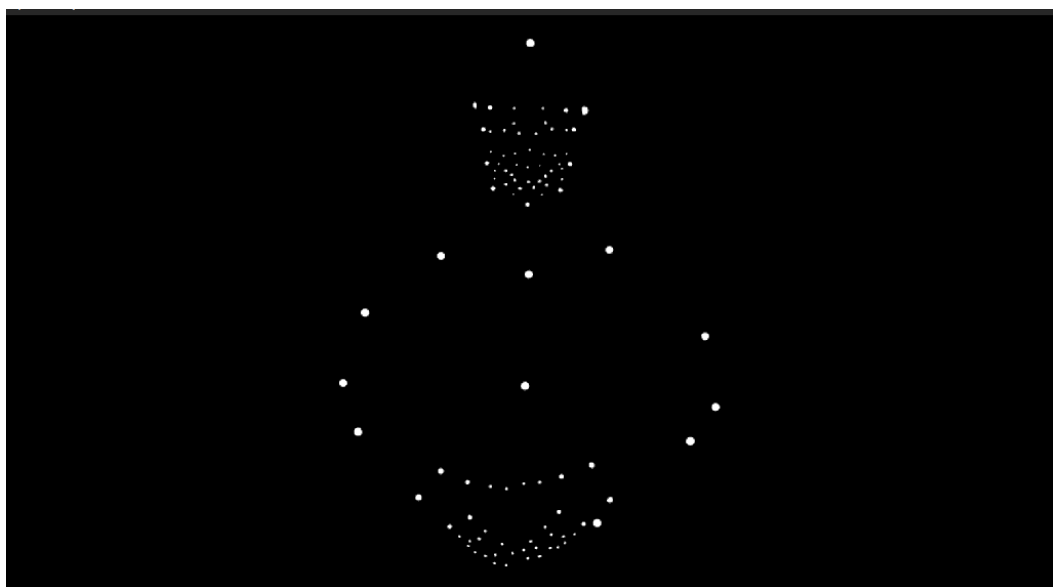
- **Optické systémy bez značek** - Nevyžadují speciální fyzické zařízení upevněné na sledovaný objekt, často se ke sledování pohybu používají barevné části oděvu. Pohyb je snímán pouze RGB kamerami. Systém se pak snaží v obraze automaticky identifikovat objekt a proložit ho kostrou.
- **Optické systémy používající značky** - Pohyb objektu je zaznamenáván více kamerami pomocí značek (dále mezinárodně používaným pojmem marker). Pro tento typ bývají využívány RGB a infračervené kamery. Systém se snaží umístit centroid do středu markeru v obraze a z polohy těchto centroidů vypočte polohu jednotlivých markerů. Tento přístup se dále dělí podle typu markerů:
  - **Aktivní** - Tyto markery udávají svojí pozici pomocí rozsvícené LED. Pro jednodušší identifikaci markeru mohou mít odlišné zbarvení nebo jich může být na jednom markeru více. Musjí být tedy napájeny z baterie či přívodním kabelem.
  - **Pasivní** - Nemají žádnou aktivní složku. Pro přístup využívající infračervené kamery se nejčastěji používají markery s retroreflexní vrstvou, která způsobuje odraz světla. Kamery v průběhu nahrávání svítí infračerveným světlem a snímají jeho odraz od markerů. Velkou výhodou tohoto systému je, že nahrávaný objekt není vázaný na žádné kabely. Tento typ systému byl použit pro získání dat, které tato práce využívá. Je proto v dalších kapitolách řeč o tomto typu. Na Obrázku 13 je znázorněn pohled jedné z kamer. Přístup využívající



RGB kamery používá markery, které mají odlišnou barvu od snímaného objektu. Tato metoda se používá např. pro zaznamenání pohybu obličeje, kdy se markery na obličej nalepí či nakreslí.

Tato metoda často bývá dražší než ostatní metody a je náchylná na chyby způsobené negativními vlivy prostředí. Mezi ně patří například odraz světla od materiálu s podobnými vlastnostmi jako marker, díky němuž se v záznamu objeví šum, který může být chybně interpretován jako detekce markeru. Dále je výrazným negativním vlivem zakrytí markeru nějakou fyzickou překážkou. Aby byl systém schopen určit polohu markeru, je třeba pohledu alespoň dvou kamer, jejichž poloha a orientace jsou známy. Na druhou stranu je tato metoda velmi přesná, data nejsou tolik zašuměná a je možné nahrávat s velkou snímkovací frekvencí.

V praxi je možné využívat hybridní systémy, které používají přístup bez značek i se značkami.



Obrázek 13: Pohled infračervené kamery na markery s retroreflexní vrstvou

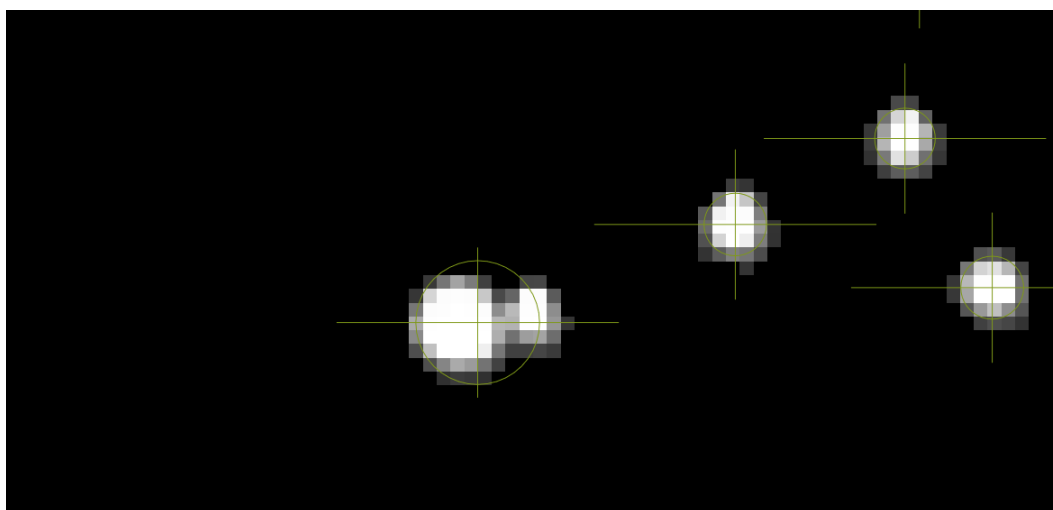
### 3.3 Postprocessing

Proces následného zpracování nahrávek nutného k získání dat se nazývá postprocessing (český výraz následné zpracování dat se příliš nepoužívá). Jednotlivé markery je třeba identifikovat a odstranit falešné markery, které se v záznamu objevily díky nechtěným vlivům. Software určený pro postprocessing dokáže vysledovat identifikovaný marker a v průběhu času si drží jeho označení. Pokud se však více markerů k sobě dostatečně přiblíží, může dojít k záměně jejich označení, což je třeba napravit.

V případě že není možné v danou chvíli vypočítat polohu markeru, je třeba jí doplnit. To je možné buď manuálním určením, nebo pomocí poloautomatické metody. Ta spočívá ve správném určení typu interpolace. Chybějící trajektorie se doplňuje například pomocí polynomické funkce dopočtené z okolních známých dat. Pro rozsáhlé mezery je však tato metoda málo přesná a používá se například kopírování trajektorie

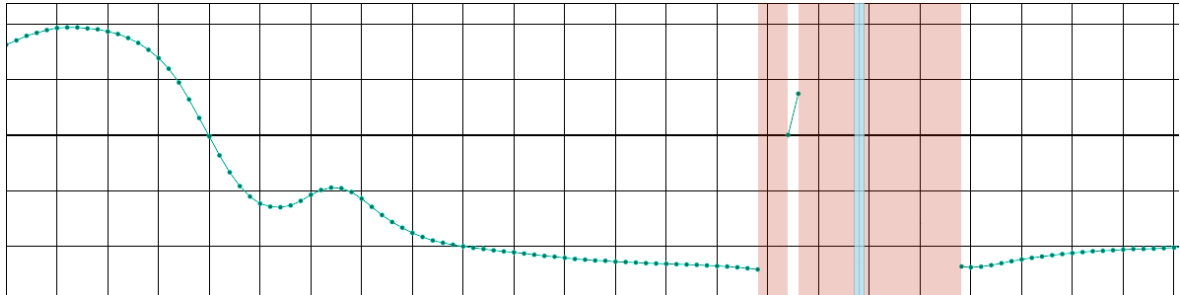
markeru, který má velmi podobný pohyb (např. markery na prstech), nebo lze vyjít z průměrné trajektorie více markerů v tuhém tělese (např. markery na pánvi nebo na hlavě). Při doplňování těchto mezer v datech je třeba vzít v úvahu jejich realističnost. Při doplňování malých mezer, kde je trajektorie pohybu zřejmá, jsou doplněná data realistická. Pokud je mezera velká tak, že trajektorii není možné realisticky doplnit, data musí být vyřazena.

Jednou z vlastností pořízených dat je šum. Ten může vznikat vícero způsoby. Zařízení není nekonečně přesné a generuje tedy po celou dobu v datech elektromagnetický šum. Další šum je generován otřesy kamer nebo jinými vlivy prostředí, jako je např. teplo, vlhkost nebo tlak. Tyto typy šumu se vyskytují v datech každého systému mocap. Díky svému stálému trvání jsou téměř neodstranitelné, mají však často v porovnání s daty malou amplitudu a nebývají tak lidským okem rozlišitelné. Dalším způsobem vzniku šumu je překrytí více markerů v obrazu všech kamer. Pokud kamery snímají markery tak, že nelze rozpoznat jejich jednotlivé centroidy, systém vyhodnotí shluk v obrazu jako jeden marker a detekuje jen jeden centroid uprostřed shluku (viz Obrázek 14). Tato detekce je tedy falešná a je nutné ji nahradit, jako by zde nebyla detekce žádná. Dále může šum vznikat vlivem nechtěného pohybu snímaného objektu, např. svalový třes. Při odstraňování takového šumu je třeba posoudit, zda jsou odšuměná data přirozenější nežli původní. Pokud šum nelze odstranit tak, aby se pohyb zdál přirozenějším, neodstraňuje se. Na Obrázku 15 je příklad doplnění chybějící trajektorie.

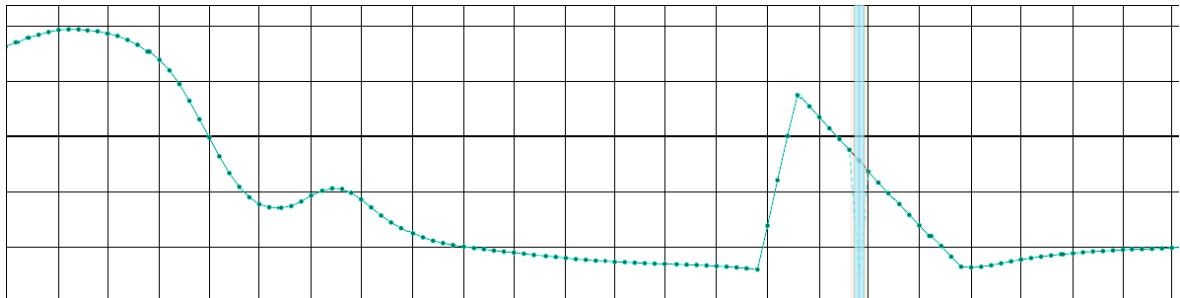


Obrázek 14: Falešná detekce markeru

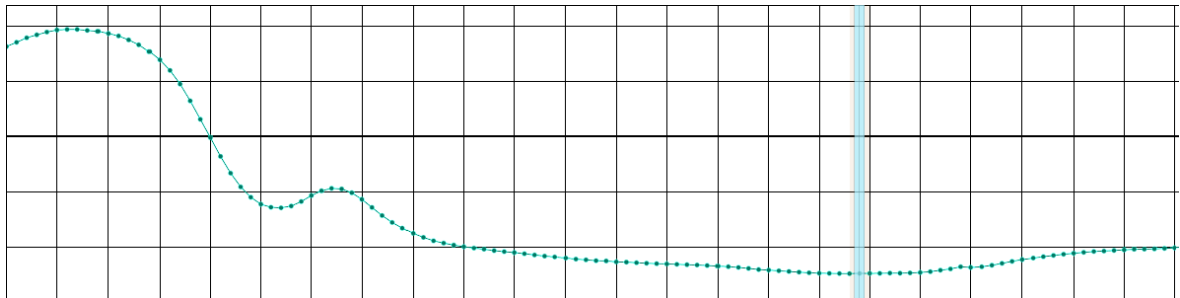
Na některých z výše uvedených metod může pracovat automat. Výslednou práci a kontrolu ale vždy musí dělat člověk. Automat nedokáže určit, který pohyb vypadá přirozeně nebo zvolit nejlepší způsob doplnění mezery. Díky tomuto faktoru bývá postprocessing časově nejnáročnější částí práce s MoCap zařízením.



(a) Originální trajektorie markeru. Červeně je označena její chybějící část, uprostřed je patrná falešná detekce nebo šum.



(b) Naivní doplnění chybějících dat interpolací.

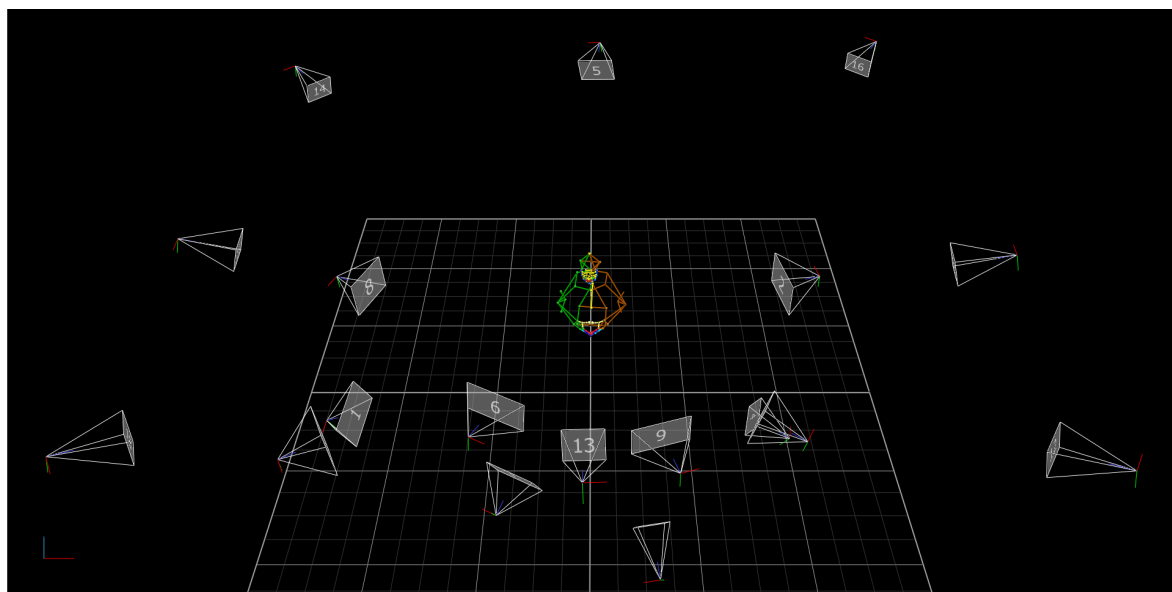


(c) Doplnění chybějících dat interpolací po odstranění šumu.

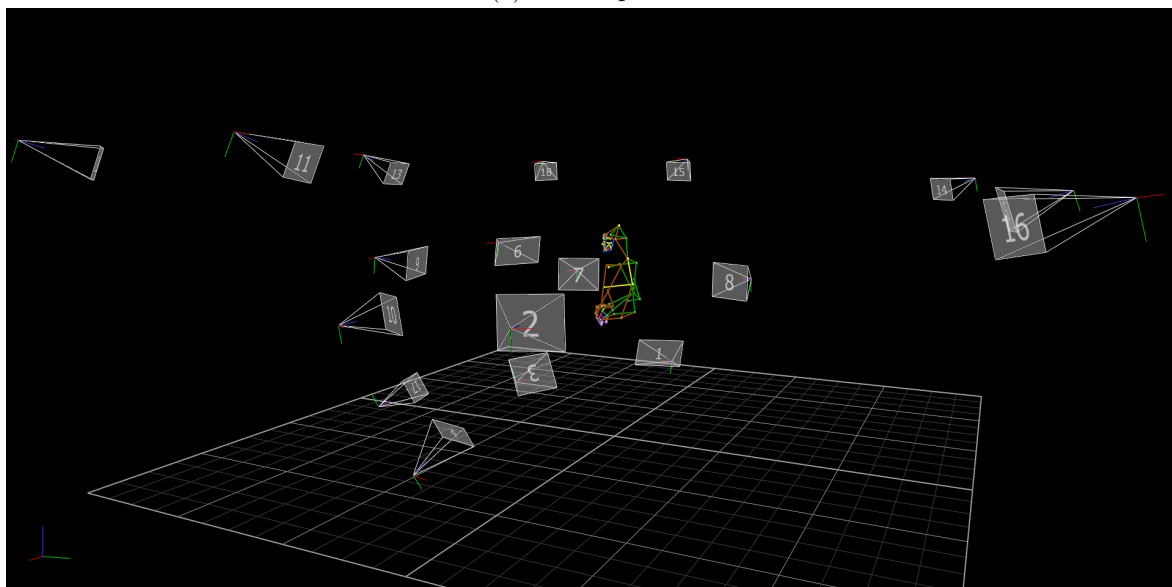
Obrázek 15: Příklad doplnění chybějících dat uprostřed trajektorie.

## 4 Data

Data pro která byla provedena analýza, byla zaznamenána systémem MoCap s celkem osmnácti kamerami. Osmi infračervenými kamerami Vicon T20 s rozlišením 2 Mpx, šesti kamerami Vicon T10 s rozlišením 1 Mpx a čtyřmi kamerami Vicon Vero s rozlišením 1,3 Mpx. Všechny tyto kamery disponují infračerveným přísvitem. Data byla pořízena se snímkovací frekvencí 120 Hz. Tato frekvence je dostatečná pro zachycení všech lidských pohybů obsažených ve znakovém jazyce. Rozmístění kamer na scéně je znázorněno na Obrázku 16.



(a) Přední pohled



(b) Boční pohled

Obrázek 16: Rozmístění kamer na scéně

## 4.1 Motivace

Neslyšící často nemohou svět vnímat sluchem. Pro prelingválně neslyšící často bývá znakový jazyk mateřským, psaný jazyk se proto učí jako cizí. v některých případech dokáží pomocí psaného textu komunikovat, je však přirozenější komunikovat ve znakovém jazyce. Proto se např. ve zprávách v televizi objevuje tlumočnick, který zprávy tlumočí do znakového jazyka. Tlumočení však kromě znalosti znakového jazyka vyžaduje mít záznam, který je následně přetlumočen, nebo v případě přímého přenosu mít schopnost tlumočit simultánně. Simultánní tlumočení bývá velmi náročné a v některých situacích, kdy je například řečník nucen zakrývat specifickou část obličeje, je obtížné řeč tlumočit. Proto je snaha vytvořit syntézu znakové řeči, tedy převod psaného textu do znakové řeči znakovanou umělým avatarem, která by mohla v některých případech tlumočnicka nahradit. Podobným způsobem se používá syntéza řeči.

Syntézu je možné tvořit více způsoby. Jejím výstupem je obraz virtuálního avatara, který tlumočí psaný vstup do znakového jazyka. Je možné syntézu založit na 2D obrazových datech a pracovat pouze s nimi. Znakový jazyk je však založen na pohybech v prostoru a obrazová data informaci o hloubce obrazu postrádají. Aby tedy byla co nejpřesnější, je pro její tvorbu využito 3D pohybových dat. Tato data jsou však oproti 2D datům velmi náročná na pořízení. Existují dva hlavní přístupy, pomocí kterých lze syntézu tvořit. Pravidlově řízená syntéza je dobře kontrolovatelná, není však dostatečně přirozená. Druhým přístupem je syntéza řízená daty. Její tvorba vyžaduje databázi předloh reálných dat s velkým počtem instancí jednotlivých znaků. Slovní zásoba je v našem případě omezena na téma předpovědi počasí. Ze serveru České televize [17], kde je k dispozici archiv zpráv ve znakovém jazyce, byl experty vybrán seznam předpovědí počasí tak, aby obsahoval znaky co nejvíce variant počasí, tedy z průběhu celého roku. v datech je tak dostatek slovní zásoby pro vytvoření nové předpovědi.

## 4.2 Pořízení dat

Pro určení polohy tuhého tělesa v prostoru stačí 3 markery. Často se jich však pro potlačení vlivu zákryvů částí těla používá více. Pokud je to možné, snažíme se ohebné těleso rozdělit po částech na tuhá tělesa tak, abychom plně zaznamenali pohyby celého tělesa. Lidské tělo lze rozdělit na tuhá tělesa (model kostry). Pohyb jednotlivých kostí se měří nepřímo umístěním markerů na kůži nebo oděv. Vzájemný pohyb kůže (oděvu) vůči kostem způsobuje neodstranitelnou systematickou chybu měření. Tato chyba je považována za zanedbatelnou.

Pro co nejpřesnější data je třeba umístit alespoň 3 markery na každou kost. Jejich umístěním do osy otáčení kloubů je možné je sdílet mezi více kostmi. Dále je počet markerů redukován, aby fyzicky nebránily přirozenému pohybu (např. na prstech). Poloha kosti tím sice není dána jednoznačně, neovlivní to však přesnost dat tolik, jako kdyby markery pohybu překázely. Na obličej byly markery umístěny tak, aby zachycovaly co nejlépe pohyb mimických svalů. Na rodilého mluvčího byly umístěny markery podle Obrázků 17 a 18 a Tabulky 1. Celkem tedy bylo umístěno 111 markerů.

## 4.3 Struktura dat

Nahraná data se dělí do dvou skupin. Skupina plynulých promluv obsahuje záznamy pohybů jednotlivých předpovědí počasí o délce průměrně 27 sekund. Skupina slovníkových dat obsahuje slovní zásobu použité znakové řeči na téma počasí. Tato slovníková

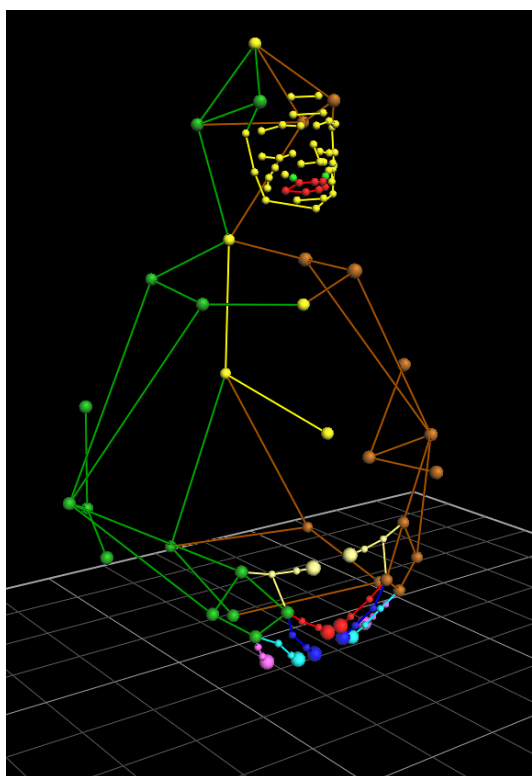


(a) Celé tělo

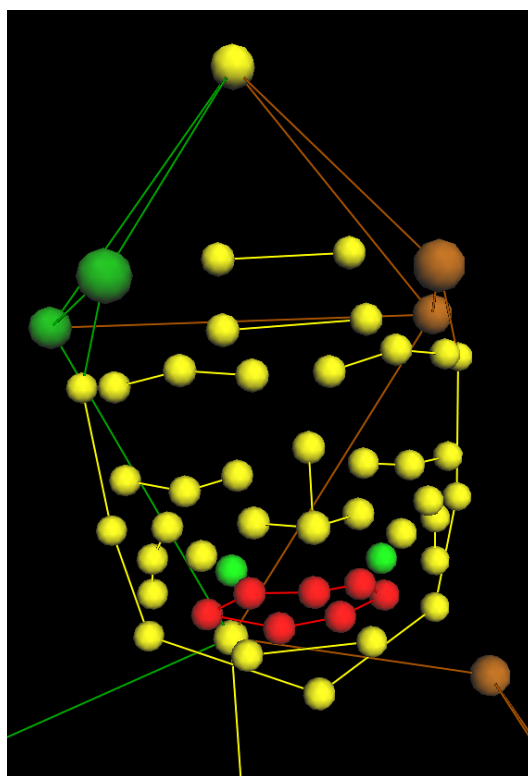


(b) Detail hlavy

Obrázek 17: Rozmístění markerů



(a) Celé tělo



(b) Detail hlavy

Obrázek 18: Vizualizace rozmístění markerů

Umístění	Počet	Průměr [mm]	Tvar
Pas	4	14	koule
Tělo	4	14	koule
Hlava	5	14	koule
Ramena	2x2	14	koule
Paže	2x4	14	koule
Zápěstí	2x2	8	koule
	2x2	6,5	koule
Prsty	2x14	4	polokoule
	2x1	6,5	koule
Obličej	35	4	polokoule
	9	6,5	koule
	4	2,5	polokoule

Tabulka 1: Přehled použitých markerů

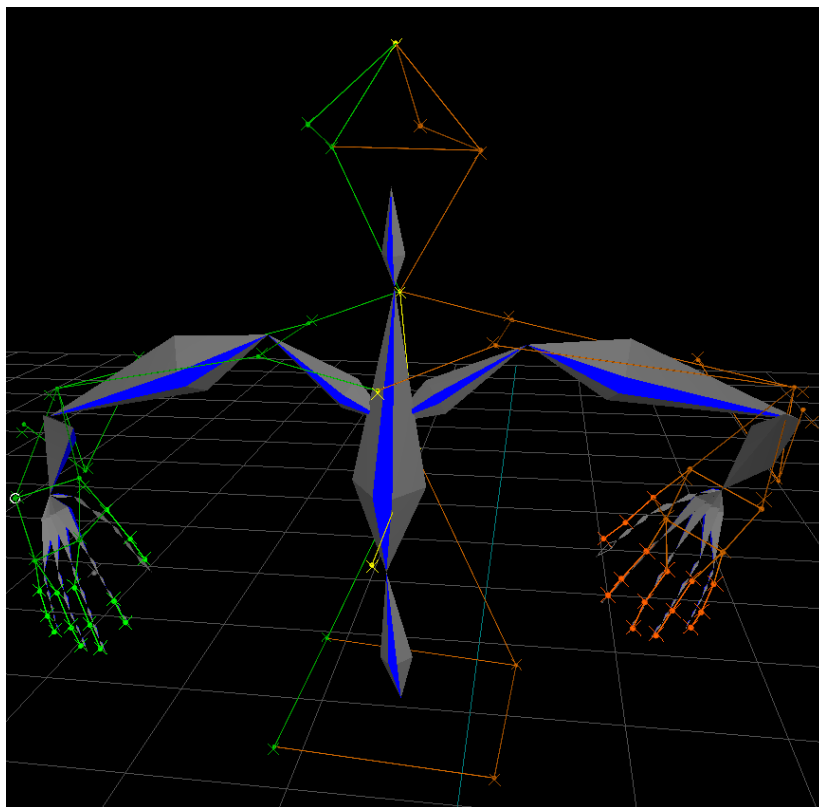
data obsahují i znaky, které nejsou v plynulých promluvách využity. Mezi jednotlivými znaky ve slovníkových datech jsou tzv. rest-pózy, kdy má mluvčí ruce spuštěné k pasu (Obrázky 17 a 18). Obě tyto skupiny jsou vhodné analyzovat zvlášť. Jedna obsahuje znaky, které jsou ovlivňovány kontextem, znaky v okolí, a druhá obsahuje každý znak zvlášť.

K pořízení záznamu pohybu byl použit systém Vicon Nexus. Software nahrává záznam z kamer a z obrazu vypočítává souřadnice jednotlivých markerů. Tato data jsou uložena v binárním souboru typu *c3d*. v souborech typu *xcp* jsou uloženy kalibrační informace kamer a ve *vsf* souborech nastavení označení markerů. Tyto soubory jsou potřebné k rekonstrukci záznamu v prostoru.

Po úpravě dat postprocessingem byla data exportována ve formátu *trc*. Tento typ souboru uchovává absolutní pozici markerů ve zvolených jednotkách s jejich označením. Dále uchovává informace o snímkovací frekvenci, délce záznamu, počtu markerů a jednotkách, ve kterých jsou data uváděna. Díky vysoké přesnosti použitého MoCap zařízení je tedy soubor *trc* nejpřesnějším způsobem uchování zaznamenaných dat. v případě editace dat v tomto formátu se však stane, že se pozice markerů změní tak, že již nebude odpovídat původnímu tělu. Aby se tomuto zabránilo, je třeba proložit tato data modelem kostry (viz Obrázek 19). Tento model není identický s lidskou kostrou, udržuje však konstantní délky částí těla. v této struktuře uchovává data soubor ve formátu *bvh*. Topologie modelu kostry je definována jako strom, kde jeho kořen odpovídá kyčelnímu kloubu (viz Obrázek 20). Tento kloub je definován vzdáleností od globální nuly. Každý další kloub je poté definován vzdáleností od svého předchůdce. Tato vzdálenost je po dobu pohybu konstantní a délky kostry tak zůstávají neměnné. Pohyb je pak definovaný rotací v jednotlivých kloubech ve stupních. Výsledky této práce jsou tedy založené právě na tomto formátu.

#### 4.4 Práce s daty

Aby bylo možné s daty pracovat, bylo třeba anotovat jednotlivé znaky v promluvách. Anotovat je třeba skutečné začátky a konce znaku, tedy i přechody mezi nimi. K přechodům je přístupováno jinak než ke znakům samotným. Protože bylo třeba data anotovat co nejpřesněji, byla anotována rodilým mluvčím znakového ja-



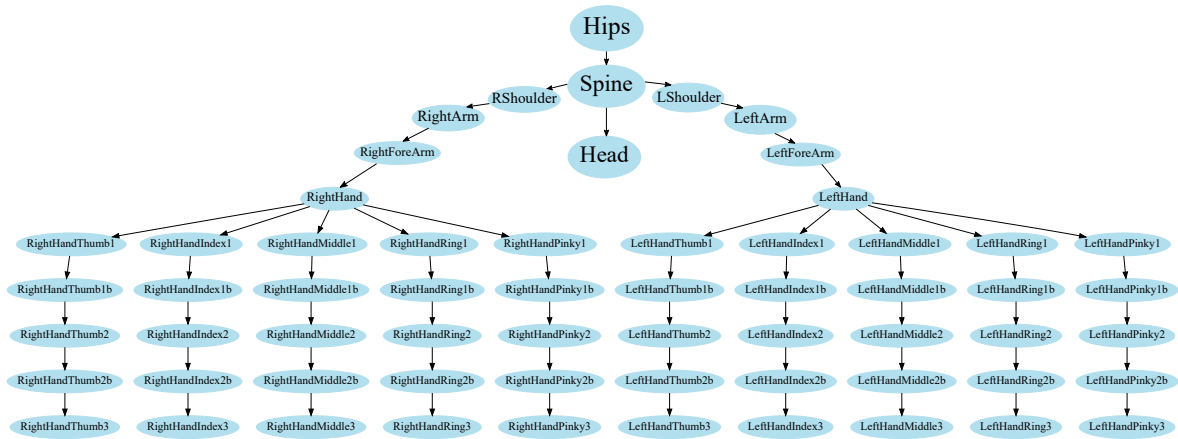
Obrázek 19: Proložení markerů (TRC) modelem kostry (BVH)

zyka. Anotátor anotoval data podle videozáznamu z nahrávání. Protože infračervené kamery pracují s frekvencí 120 Hz a RGB kamera použitá k pořízení videozáznamu pouze 24 Hz, dochází k drobným nepřesnostem.

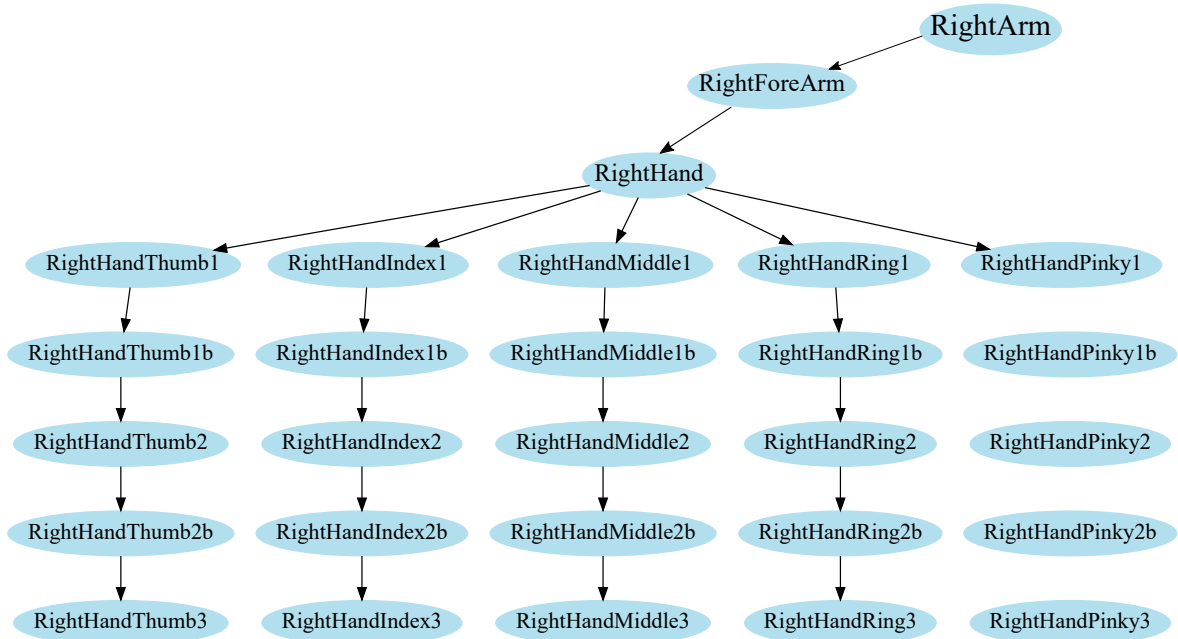
Syntéza znakového jazyka bude vytvořena pomocí strojového učení. To však vyžaduje velké množství vstupních dat. Jedním ze způsobů, kterým lze množství dat navýšit je augmentace. Augmentací je nazýván proces, kdy se data upravují tak, aby stále byla realistická a bylo možné použít jako nová vstupní data.

Pohyb nemanuální složky znakového jazyka byl zaznamenán. Protože však obecně nese jinou informaci než manuální složka, je tato analýza zaměřena pouze na manuální složku. Pro nemanuální složku by byla třeba udělat samostatná analýza, avšak důraz je kladen nejdříve na manuální složku. v případě práce s nemanuální složkou, především specifickými znaky (viz 2.2.2), je třeba vzít v úvahu jev zvaný foreign talk.





(a) Struktura celého těla



(b) Detail struktury pravé ruky

Obrázek 20: Stromová struktura použitých *bvh* dat

## 5 Analýza

Existuje více způsobů, jakými lze data augmentovat. Mezi ně patří zašumění reálných dat, jejich translace a rotace nebo mutace s jinými instancemi znaků. Aby bylo možné určit, do jaké míry lze data upravovat tak, aby odpovídala reálným, je třeba analyzovat data a vybrat měřitelné veličiny, které k tomu mohou přispět. Cílem této práce je vybrat veličiny a metody, které lze pro daná data měřit a přispět tak k tvorbě umělých dat ve smyslu jejich tvorby i zhodnocení jejich věrohodnosti. Toto hodnocení může být pouze orientační. Automat nikdy nedokáže lépe určit přirozenost pohybu lidského těla než člověk. Proto finální zhodnocení kvality augmentace bude vizuální kontrolou člověkem.

### 5.1 Pozice

Jak již bylo zmíněno v kapitole 4.3, používaný formát dat ukládá informace o natočení jednotlivých kloubů vůči předchozím. Díky této reprezentaci dat je zachována struktura lidského těla. Není však zřejmá informace o pozici jednotlivých částí těla mezi sebou, o jejich kontaktu. Intuitivnějším způsobem zápisu jsou zpětně vypočtené absolutní polohy z angulárního formátu. Jedná se o stejná data, která lze navzájem přepočítat. Pro některé z níže popsaných metod však předzpracování touto metodou může být vhodné.

Z informace o absolutní pozici jednotlivých částí těla lze určit v jaké poloze jsou ruce oproti tělu nebo hlavě. HamNoSys (viz 2.3.3) pro zápis manuální složky používá tvar ruky, orientaci, lokaci a pohyb. Díky výpočtu absolutní pozice všech kloubů, lze rozdělit prostor kolem těla do úseků podle HamNoSys a určit lokaci znaků. Dále lze určit vzdálenost jednotlivých kloubů od sebe. Toho lze využít k určení vztahu rukou v pohybu znaku.

### 5.2 Rychlost a zrychlení

Základní metrikou, kterou lze pro naměřená data získat, je okamžitá rychlost (1) a zrychlení (2). Lidské tělo má omezenou maximální rychlost a zrychlení, kterými se může pohybovat. Při tvorbě augmentovaných dat je tedy vhodné počítat maximální rychlost a zrychlení jednotlivých promluv. Pokud uměle vytvořený pohyb dosáhne určitých hodnot, jedná se o nepřirozený pohyb. Tento případ může nastávat například při tvorbě syntézy pokud je zvolena délka přesunu mezi jednotlivými znaky moc malá. Tělo bude muset vyvinout velkou rychlost aby se přesunulo z jedné pozice do druhé během krátké doby.

$$v = \frac{d}{dt} \sqrt{x^2 + y^2 + z^2} \quad (1)$$

$$a = \frac{dv}{dt} \quad (2)$$

### 5.3 Dynamika pohybu

Vektor rychlosti je možné dále využít k vypočtení průměrné vzdálenosti uražené za jeden snímek  $D$ . Veličinu získáme podle (3). Vypočtená hodnota udává dynamiku pohybu, čím je tato hodnota větší, tím je dynamika pohybu větší.  $D$  využívají některé z níže popsaných metod.

$$D = \frac{\sum_{t=1}^T |v_t|}{T} \quad (3)$$

Dynamika pohybu může sloužit k určení míry pohybu různých znaků. Jednou ze základních myšlenek augmentace znaků je mutace podobných znaků mezi sebou. Rozdíl hodnot  $D$  mutovaných znaků může být jednou z nutných podmínek pro její validitu. Přestože všechny znaky jeví známky dynamiky, lze je z pohledu datové analýzy rozdělit na znaky statické a dynamické. Znak je považován za statický, pokud nese veškerou informaci v jednom snímku, např. znaky pro čísla 1-5.

### 5.4 Dominance ruky

Pojmem lateralita bývá označováno upřednostňování jednoho z párových orgánů. Lateralita rukou se ve znakovém jazyce projevuje používáním dominantní ruky pro jednoruční znaky nebo jako aktivní ruky pro dvouruční znaky. Nedominantní ruka pak může hrát pasivní i aktivní roli v dvouručních znacích, nebo zůstává v klidu u jednoručních znaků (viz kapitola 2.2). v kapitole 5.3 je definována dynamika pohybu  $D$ , která udává s jakou mírou se tělo pohybuje. Pokud je  $D$  vypočteno pro každou ruku zvlášť, je možné je pak navzájem porovnat. Vydělením  $D$  levé a pravé ruky získáme poměr dynamiky pohybu  $r$  (4).

$$r = \frac{D_L}{D_P} \quad (4)$$

Dominantní ruka se oproti nedominantní pohybuje více a její  $D$  tak nabývá vyšších hodnot. Je tedy teoreticky možné díky tomuto poměru určit dominantní ruku. Pokud je výsledná hodnota  $r < 1$ , je dominantní pravá ruka a pokud  $r > 1$ , je dominantní levá. Pro  $r = 1$  je  $D$  obou rukou stejné. Čím blíže se tedy hodnota  $r$  pohybuje 1, tím je určení dominantní ruky méně spolehlivé.

Po celou dobu obecného pohybu znakového jazyka je podle [18] poměr  $r$  nedominantní ruky vůči dominantní  $r = 0,78$ . Benchiheub pro svůj výpočet používá data v kartézských souřadnicích a nevyužívá pohybu prstů. Pro všechny naměřené promluvy byl tento poměr vypočten jak pro klouby od ramene po zápěstí, tak i pro všechny následující klouby včetně prstů. v Tabulce 2 jsou zapsány hodnoty vypočteného poměru pro angulární data i pro data přepočtená do kartézských souřadnic.

	Včetně prstů	Pouze paže
Angulární	0,6970	0,6536
Kartézské	0,8944	0,8015

Tabulka 2: Poměr dynamiky levé ruky vůči pravé

Hodnota vypočteného poměru kartézských dat bez prstů se jen mírně liší od hodnoty v [18]. Tato odchylka je zřejmě způsobena odlišným umístěním markerů, řečníkem,

nebo promluvou, pro kterou je pohyb zaznamenáván. z Tabulky 2 lze vyčíst, že z dat v angulární reprezentaci lze dominanci určit lépe než z kartézské a že pohyb prstů v průběhu promluvy nenese dodatečnou informaci o dominanci. Pro určení dominantní ruky je tedy z testovaných metod nejvhodnější použití angulárních dat pouze pro paže. Pro všechny metody vychází  $r < 1$ , což odpovídá skutečnosti, že znakování má dominantní pravou rukou.

## 5.5 Určení jednorukosti-dvourukosti

Poměr délek trajektorie  $r$  by mohl být mimo určení dominantní ruky také vhodný k určení, zdali je znak jednoruční nebo dvouruční. Určení jednorukosti-dvourukosti nemusí vždy odpovídat lingvistickému dělení. Předpokladem pro jednoruční znaky jsou nízké hodnoty  $r$  a vysoké hodnoty pro dvouruční znaky. Prsty mají v použité struktuře mnohem více kloubů než paže. Pokud by bylo  $D$  počítáno pro všechny klouby se stejnou vahou, došlo by k výraznému potlačení pohybu paží a ke zvýraznění pohybu prstů. Tento jev lze omezit uvažováním pouze  $D$  pro ruce bez prstů. Dalším způsobem je přiřazení jednotlivým kloubům váhy. v tomto případě jsou klouby váženy podle délky kosti, tedy vzdálenosti od předchozího kloubu. Klouby na ruce jsou tak počítány s větší vahou než klouby na prstech. v budoucnu je pro další analýzu možné vážit klouby podle jiných heuristik.

Pro všechny znaky plynulých i slovníkových promluv byl pomocí metody support vector machines (SVM)<sup>13</sup> vypočten práh poměru, podle kterého lze znaky automaticky rozdělit na jednoruční a dvouruční. SVM patří mezi metody strojového učení s učitelem. Slouží ke klasifikaci vstupních dat do tříd na základě trénovacích dat, pro která je zadán požadovaný výstup. Pro určení požadovaného výstupu byl pro celou slovní zásobu anotován počet znakování rukou. Aby bylo možné otestovat správnost výsledku SVM, byla data předem náhodně rozdělena na 90 % trénovacích a 10 % testovacích dat. Pro tato data byla vypočtena úspěšnost metody SVM. Protože je výsledek závislý na náhodném výběru testovacích dat, bylo uskutečněno  $10^5$  opakování. Střední hodnota úspěšnosti byla vypočtena pro znaky plynulých i slovníkových promluv, pro klouby vážené délkou kostí i pouze pro klouby rukou a pro data v kartézské soustavě souřadnic i pro angulární data.

Pohyb rukou znaků plynulých promluv je ovlivněn znaky, které aktuálnímu znaku přecházely a budou následovat.  $D$  je tedy počítáno také pro sousední znaky (tzv. trigram). Znaky slovníkových promluv jsou navzájem odděleny klidovou pózou, proto nemá význam tvořit trigramy pro slovníková data. Pro trigramy bylo uskutečněno  $10^5$  opakování obdobných samostatným znakům (unigramům). Výsledky všech použitých metod jsou zaznamenány v Tabulce 3.

Z výsledků v Tabulce 3 je možné vidět, že žádná z metod nedosahuje úplné přesnosti. To znamená, že se v některých případech dvouručních znaků (viz Obrázek 1), které mají jednu ruku jako pasivní (viz 2.2), tato ruka pohybuje stejnou nebo menší mírou než u jednoručních znaků. Naopak u jednoručních znaků se může nedominantní ruka hýbat stejně nebo více než dominantní, aniž by nesla nějakou informaci, například při přesunu na lokaci následujícího znaku. Díky negativním vlivům okolních znaků dosahuje metoda horších výsledků pro plynulé promluvy než pro slovníková data, která okolními znaky ovlivněná nejsou. Dále lze z tabulky vidět, že pro slovníková data je vhodnější reprezentace kartézskými souřadnicemi, zatímco pro plynulé promluvy je

<sup>13</sup>Pro výpočet byla použita knihovna sklearn v0.22.2 (pro Python3.7)

			Angulární souřadnice		Kartézské souřadnice	
			Vážené	Pouze paže	Vážené	Pouze paže
Unigramy	Slovníková data	Stř.hodnota	0,9630	0,9516	0,9909	0,9387
		Variance	0,0010	0,0013	0,0002	0,0016
	Plynulé promluvy	Stř.hodnota	0,8568	0,8807	0,7868	0,7888
		Variance	0,0012	0,0010	0,0016	0,0016
Trigramy	Plynulé promluvy	Stř.hodnota	0,8737	0,8848	0,7701	0,7768
		Variance	0,0013	0,0012	0,0022	0,0022

Tabulka 3: Úspěšnost metody SVM pro určení jednorukosti-dvourukosti

vhodnější angulární reprezentace. Metoda využívající trigramy nemá oproti metodě pracující s unigramy výrazně odlišné výsledky pro žádný způsob reprezentace dat. Trigramy tedy nejsou reprezentací dat, která by výrazně přispěla k lepším výsledkům této metody. z výsledků je zřejmé, že úspěšnost rozpoznání jednorukosti-dvourukosti plynulých promluv pouze podle hodnoty poměru délek trajektorií není dostatečná. Pro tuto úlohu tedy není uvedena metoda vhodná. Při tvorbě umělých dat je však možné kontrolovat meze v jakých se poměr  $r$  pro daný znak pohybuje a použít ji tak jako jednu z hodnotících veličin.

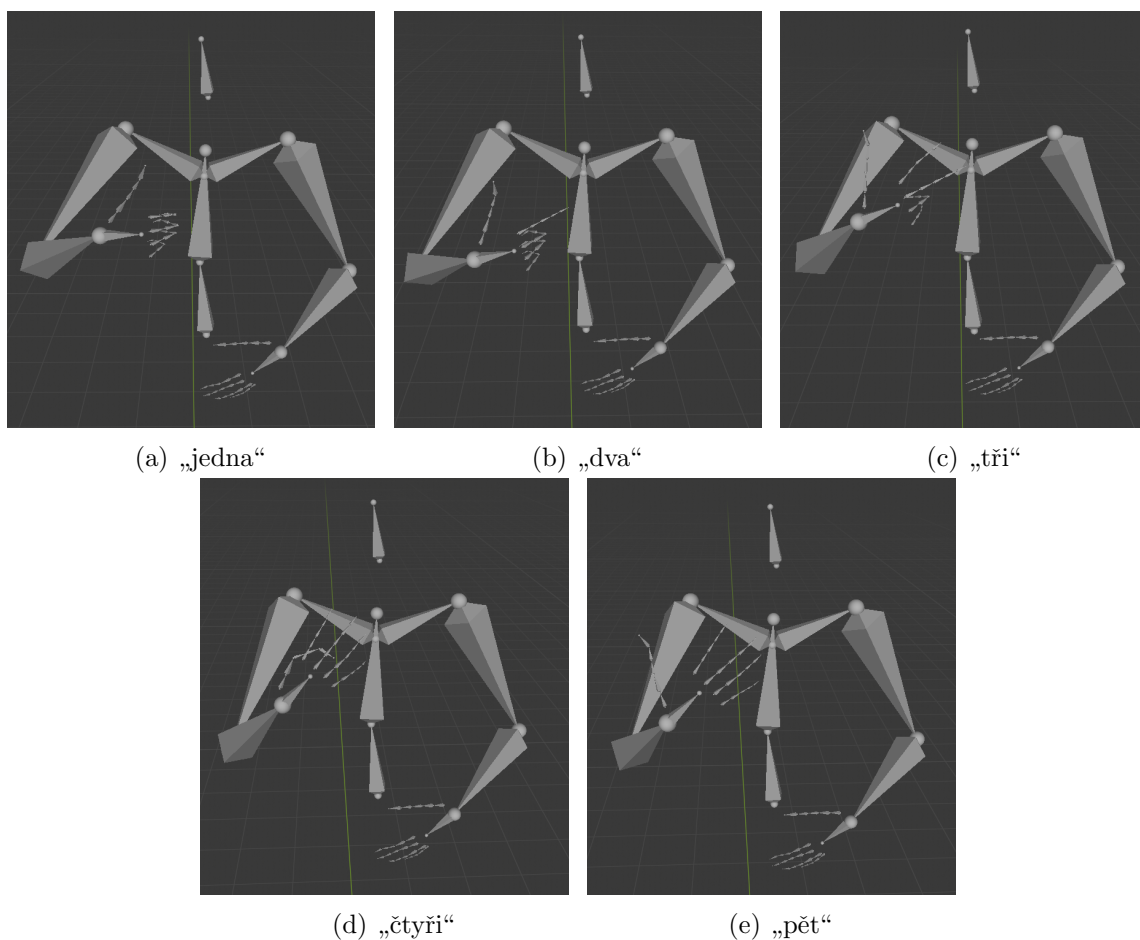
## 5.6 Vyhodnocení podobnosti

Jediným vyhodnocením přirozenosti umělých dat je hodnocení člověkem. Aby hodnocení bylo co nejobjektivnější, je třeba, aby data hodnotilo dostatečné množství lidí, což je časově náročné. Stroj nedokáže přirozenost pohybu určit, dokáže však určit podobnost dat navzájem. Pokud se umělá data dostatečně podobají datům, která byla získána nahráváním skutečného řečníka, lze jejich pohyby pokládat za přirozené. Základní metrikou, kterou lze využít k určení podobnosti dat, je euklidovská vzdálenost. Vztah pro výpočet euklidovské vzdálenosti mezi vektory  $X$  a  $Y$  je (5). Čím je výsledná hodnota menší, tím jsou si vektory podobnější. Euklidovská vzdálenost je počítána pro vektory schodných rozměrů. Lze ji proto vypočítat pouze pro dva znaky schodné délky.

$$d_{X,Y} = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2} \quad (5)$$

Ze slovní zásoby byly vybrány znaky, které mohou nést informaci pouze v jednom snímku, tudíž je vzdálenost počítána pro vektory o rozměru rovném počtu kloubů. Významy těchto znaků jsou „jedna“, „dva“, „tři“, „čtyři“ a „pět“. Jejich vizualizace je na Obrázku 21. Tyto znaky byly vybrány také proto, že se liší pouze ve tvaru ruky, pozici a držení těla mají stejné. Pro všechny zastoupení (v plynulých promluvách i slovníkových) vybraných znaků navzájem byla vypočítána euklidovská vzdálenost. Výsledky jsou zapsány v Tabulce 4. Porovnány jsou vzdálenosti počítány pro klouby celého těla a pouze pro klouby rukou v angulární reprezentaci.

Euklidovskou vzdálenost je možné počítat pouze pro znaky stejné délky. I významově stejné znaky však často mají rozdílnou délku provedení. Proto je vhodné ji počítat pouze pro statické znaky[5.3], které nesou svojí informační hodnotu jedním snímekem. Pro výpočet podobnosti různě dlouhých znaků je vhodné využít jiných prostředků.

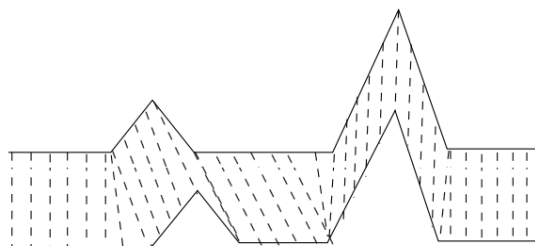


Obrázek 21: Vizualizace znaků pro čísla 1-5 na použitém modelu kostry

Znak	Poč.	Vzdálenost pro všechny klouby		Vzdálenost pro klouby na ruce	
		Stejný význam	Rozdílný význam	Stejný význam	Rozdílný význam
„jedna“	3x	89,35 - 166,92	146,64 - 344,84	47,21 - 72,43	125,99 - 285,99
„dva“	5x	49,79 - 214,25	146,64 - 324,96	11,19 - 39,31	116,87 - 259,12
„tři“	2x	135,67 - 135,67	160,98 - 255,62	36,2 - 36,2	128,23 - 218,73
„čtyři“	3x	52,58 - 206,63	194,97 - 354,6	8,6 - 34,81	146,8 - 287,8
„pět“	3x	60,43 - 102,89	149,6 - 304,49	12,16 - 30,18	130,2 - 277,72

Tabulka 4: Euklidovská vzdálenost

Jedním z nich je funkce Dynamic time warping (DTW)<sup>14</sup>. Jedná se o metodu porovnání dvou vektorů, které mohou být rozdílné délky. Algoritmus prohledává možné způsoby zarovnání vektorů tak, aby jejich vzdálenost byla co nejmenší. Na Obrázku 22 je znázorněno zarovnání dvou vektorů. Díky možnosti porovnání rozdílně dlouhých vektorů je možné porovnávat znaky, jejichž rychlost se v čase liší. Čím je hodnota menší, tím si jsou znaky podobnější. Pro stejné znaky, jako byla počítána euklidovská vzdálenost, byla vytvořena obdobná tabulka s výsledky DTW funkce (Tabulka 5). Znaky však nyní nejsou reprezentovány pouze statickou pózou, ale sekvencí těchto póz určených anotátorem. Výsledky DTW funkce byly normalizovány vydělením délkou nalezené cesty a počtem kloubů použitých pro výpočet.



Obrázek 22: Dynamic time warping<sup>15</sup>

Znak	Poč.	DTW pro všechny klouby		DTW pouze pro klouby rukou	
		Stejný význam	Rozdílný význam	Stejný význam	Rozdílný význam
„jedna“	3x	1,845 - 3,0457	2,1799 - 6,0704	1,9698 - 3,2517	2,3274 - 6,4811
„dva“	5x	0,973 - 3,3638	2,1799 - 5,5986	1,0388 - 3,5914	2,3274 - 5,9774
„tři“	2x	2,4168	2,4492 - 4,4066	2,5803	2,6148 - 4,7047
„čtyři“	3x	0,8331 - 3,0733	3,3959 - 6,4686	0,8895 - 3,2812	3,6257 - 6,9062
„pět“	3x	1,0332 - 1,9416	2,4492 - 4,9455	1,1031 - 2,0729	2,6148 - 5,2800

Tabulka 5: Dynamic time warping

Dalším způsobem, kterým lze určit podobnost znaků rozdílných délek, je výpočet korelační matice. Korelace je vztah vyjadřující míru lineární závislosti dvou veličin. Pro matice  $A$  a  $B$  je korelační matice  $P$  počítána dle (6). Koeficienty korelační matice mohou nabývat hodnot  $\rho \in \langle -1; +1 \rangle$ . Pokud je  $\rho = 0$ , veličiny jsou na sobě lineárně nezávislé, pokud  $|\rho| = 1$ , jsou na sobě lineárně závislé. Pro matice časových průběhů znaků byly vypočteny korelační matice. Průměry jejich prvků jsou zapsány v Tabulce 6.

$$P = \frac{\sum_m \sum_n (A_{mn} - \bar{A}) (B_{mn} - \bar{B})}{\sqrt{\left( \sum_m \sum_n (A_{mn} - \bar{A})^2 \right) \left( \sum_m \sum_n (B_{mn} - \bar{B})^2 \right)}}, \quad (6)$$

kde  $\bar{A}$  je střední hodnota  $A$  a  $\bar{B}$  je střední hodnota  $B$

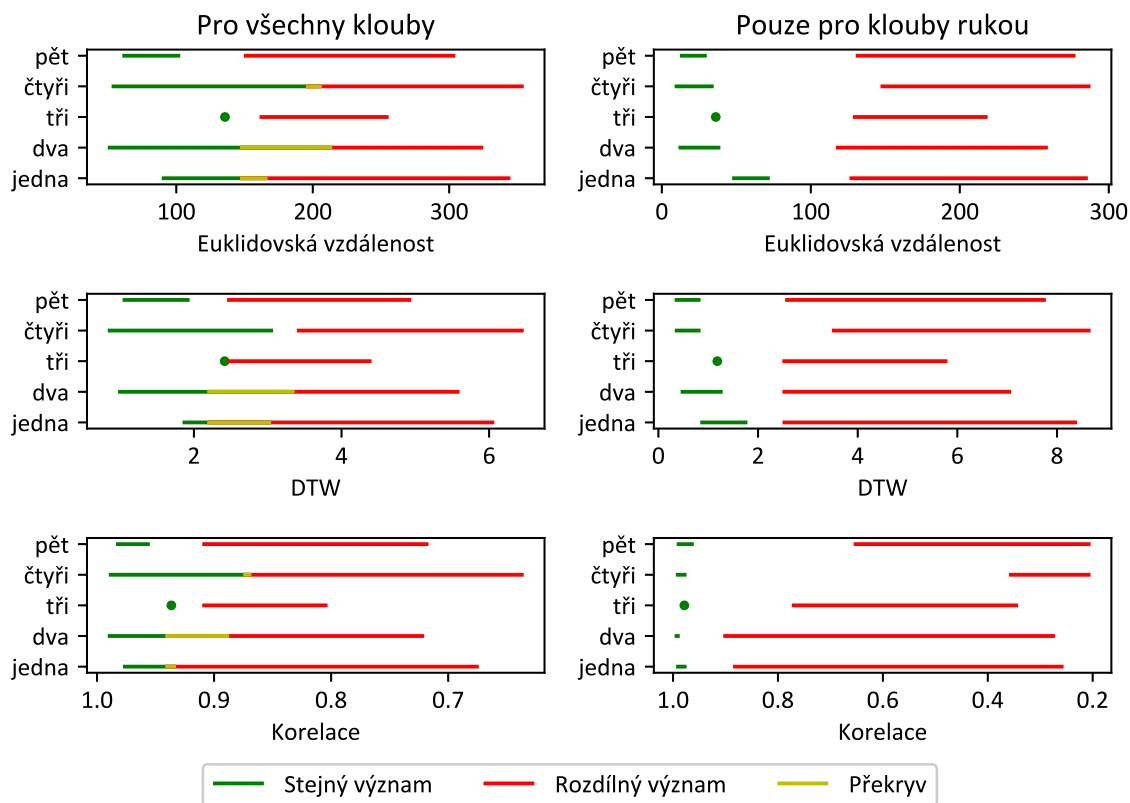
<sup>14</sup>Pro výpočet byla použita knihovna fastdtw v0.3.4 (pro Python3.7)

<sup>15</sup>Zdroj: [https://commons.wikimedia.org/wiki/File:Dynamic\\_time\\_warping.png](https://commons.wikimedia.org/wiki/File:Dynamic_time_warping.png)

Znak	Poč.	Korelace pro všechny klouby		Korelace pouze pro klouby rukou	
		Stejný význam	Rozdílný význam	Stejný význam	Rozdílný význam
„jedna“	3x	0,9742 - 0,9942	0,2555 - 0,8856	0,9325 - 0,9778	0,6739 - 0,9417
„dva“	5x	0,9873 - 0,9970	0,2716 - 0,9043	0,8872 - 0,9908	0,7205 - 0,9417
„tři“	2x	0,9785	0,3422 - 0,7735	0,9365	0,8031 - 0,9101
„čtyři“	3x	0,9742 - 0,9947	0,2040 - 0,3595	0,8681 - 0,9899	0,6355 - 0,8751
„pět“	3x	0,9605 - 0,9930	0,2040 - 0,6557	0,9549 - 0,9838	0,7168 - 0,9101

Tabulka 6: Korelace

Výsledky experimentů jsou vizualizovány na Obrázku 23. z grafů lze vidět, že všechny metody lépe dokážou rozlišit znak podle tvaru ruky nežli podle celého těla. Horší rozlišitelnost celého těla je způsobena rozdílným držení těla, hlavy a paží, zatímco je jejich poloha teoreticky stejná. Tento jev by bylo možné potlačit např. heuristickým vážením jednotlivých kloubů podle hodnoty nesené informace. Dále lze z grafů vidět, že pro veličiny počítané pouze pro klouby rukou jsou znaky stejného a rozdílného významu nejlépe rozlišitelné pomocí metody korelace. Tyto výsledky jsou však velmi ovlivněné malým počtem dat, výběrem testované množiny znaků a nebo také např. řečníkem.



Obrázek 23: Vizualizace experimentů určení podobnosti

Aby bylo možné porovnat výše zmíněné metody s člověkem, byla vytvořena jednoduchá augmentace statických znaků čísel 1-5. Byly vytvořeny všechny kombinace těchto znaků záměnou tvarů rukou. Sedm dobrovolníků mělo za úkol ohodnotit přirozenost



držení těla k danému tvaru ruky na Likertově škále hodnocením 1-5 (1-nejlepší, 5-nejhorší). Výsledky tohoto experimentu jsou zaznamenány v Tabulce 7. z Obrázku 24 lze pozorovat, že člověk není schopný jednoznačně určit skutečná data. Toto může být způsobeno tím, že vizualizace umělých dat byla na modelu kostry (viz Obrázek 21) a člověk nehumanoidní vizualizace vnímá jako nepřirozené. Dále lze pozorovat, že vyhodnocení má jistou varianci, ve které člověk hodnotí přirozenost. Příkladem je znak pro číslo 2, kde člověk hodnotí skutečný znak hůře než znaky jemu vzdálené (viz Obrázek 24b). Výsledky tohoto experimentu mohou být závislé nejen na výběru použitých dat a na typu augmentace, ale také na počtu lidských hodnocení.

		Tvar ruky				
		„jedna“	„dva“	„tři“	„čtyři“	„pět“
Držení těla	„jedna“	1,7	2,9	2,9	3,9	3,9
	„dva“	1,1	2,6	2,7	3,1	3,4
	„tři“	1,7	1,6	2	2,1	2,285714
	„čtyři“	3	1,4	1,1	1	1
	„pět“	3	1,6	1,4	1,142857	1,142857

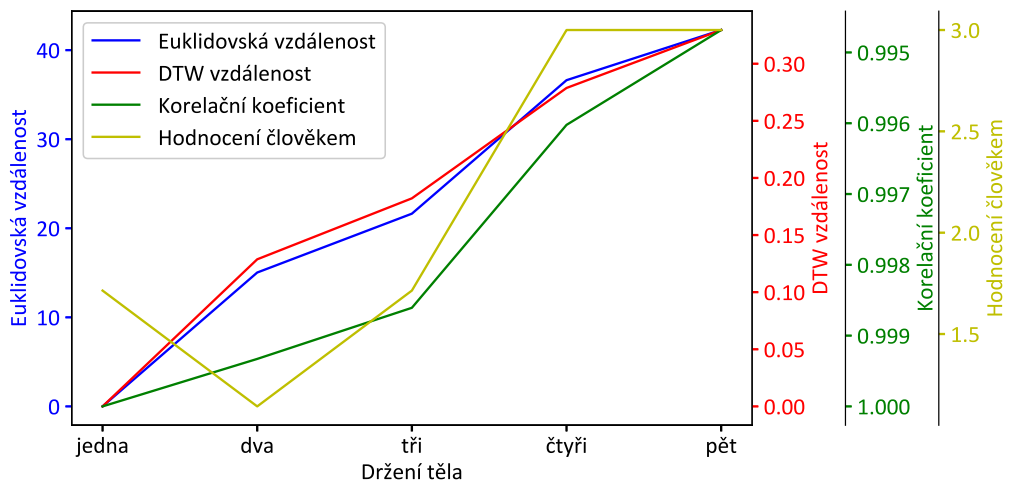
Tabulka 7: Hodnocení umělých dat člověkem

## 5.7 Symetrie znaků

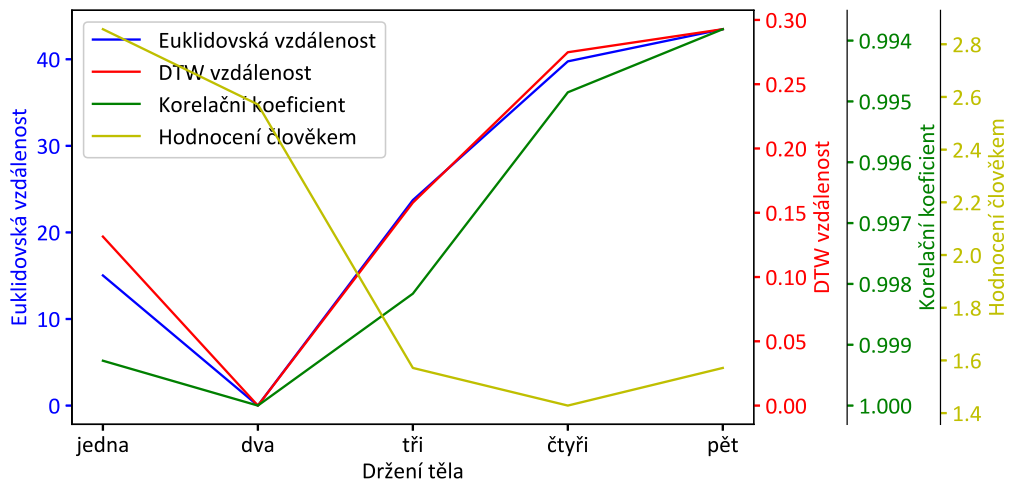
Podle způsobu artikulace se dvouruční znaky dají z pohledu symetrie rozdělit na 3 skupiny. Znaky artikulované aktivně oběma rukama se stejným tvarem. Symetrické znaky, které mají jednu ruku aktivní a druhou pasivní, ale obě mají stejný tvar ruky. A asymetrické, které mají také jednu ruku aktivní a druhou pasivní, ale tvary rukou mají rozdílné [4]. Míra symetrie z lingvistického hlediska je řešitelná pomocí metod diskutovaných v kapitole 5.6, kdy se měří podobnost tvaru levé a pravé ruky. Tato kapitola však nepojednává o symetrii z lingvistického hlediska, ale za symetrické znaky jsou považovány ty, které jsou zrcadlově souměrné.

Cílem úlohy je automaticky určit zdali je znak symetrický pomocí výpočtu vzájemné korelace mezi rukama. Pro jednotlivé klouby paží (rameno, loket a zápěstí) byl vypočten korelační koeficient podle (7) pro každou dimenzi zvlášť. Pro metodu SVM, která již byla použita v kapitole 5.5, byla data rozdělena na 90 % trénovacích dat a 10 % testovacích. Jako požadovaný výstup byl použit vektor předem anotovaných informací o symetrii jednotlivých znaků. Protože výsledek je závislý na náhodném výběru testovacích dat, byl tento výpočet realizován  $10^5$  krát. Výsledky této metody pro data v angulárních i kartézských souřadnicích jsou zapsány v Tabulce 8. z tabulky lze vidět, že pro určení symetrie znaků je vhodnější angulární reprezentace dat.

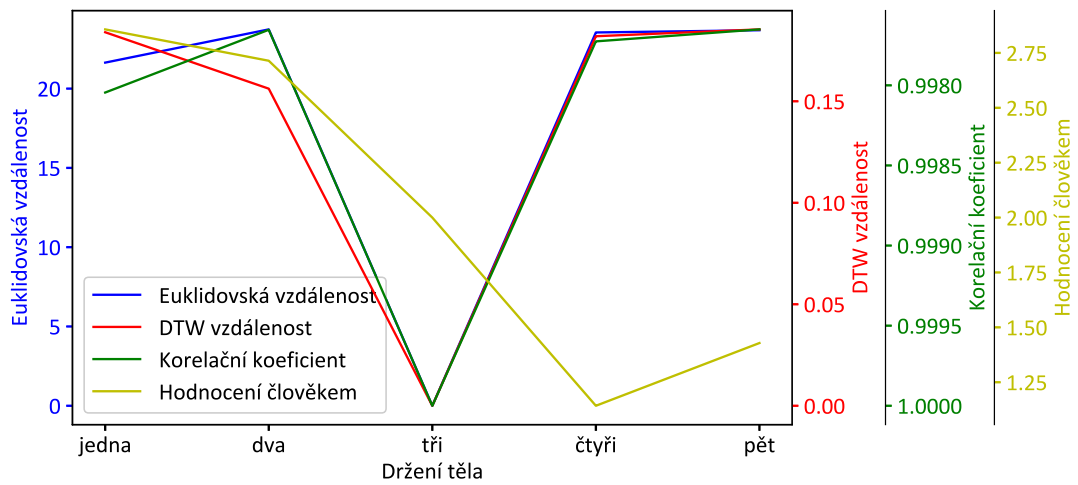
$$\rho_{X,Y} = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - E(X)^2} \cdot \sqrt{E(Y^2) - E(Y)^2}} \quad (7)$$



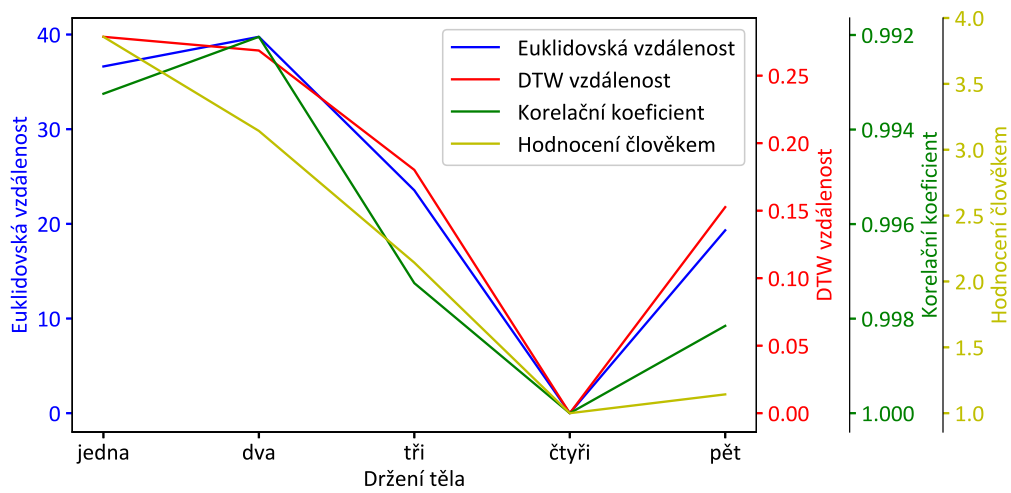
(a)



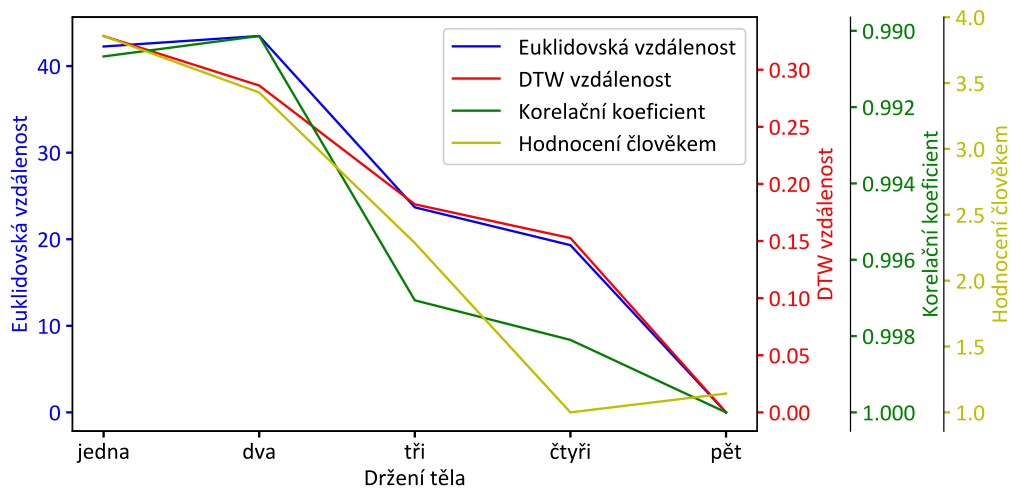
(b)



(c)



(d)



(e)

Obrázek 24: Na grafech 24a, 24b, 24c, 24d a 24e jsou výsledky použitých metod pro znaky „jedna“, „dva“, „tři“, „čtyři“ a „pět“

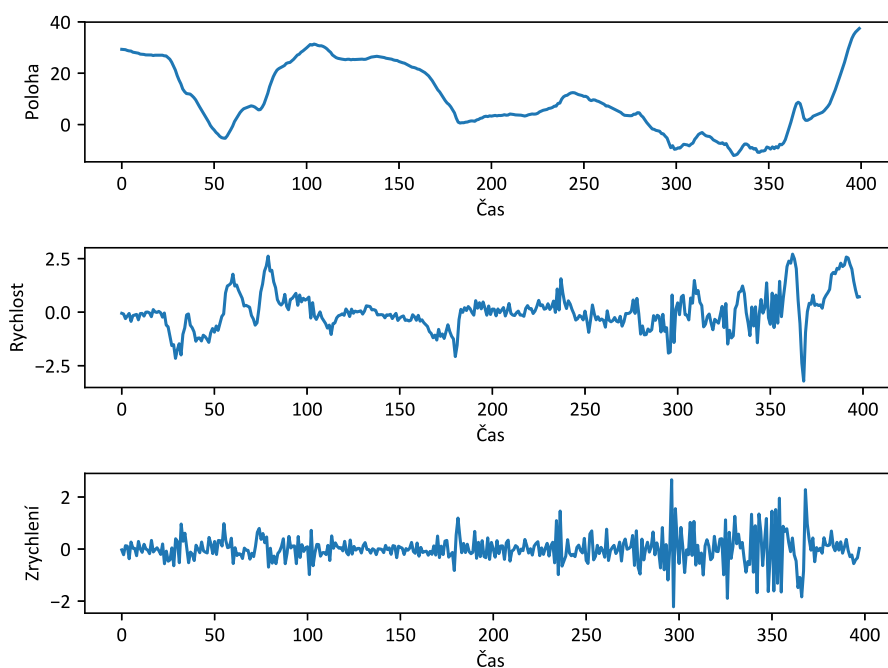
	Střední hodnota	Variance
Angulární	0,9125	0,0021
Kartézské	0,8652	0,0032

Tabulka 8: Úspěšnost metody SVM pro určení symetrických znaků

## 5.8 Frekvenční analýza

Data znakové řeči je možné segmentovat v rámci celých promluv i jednotlivých znaků. Typů segmentace může být celá řada. Příkladem jednoho z jejich výsledků je nalezení klidových póz či hraniční snímky znaků. Tuto informaci sice obsahuje anotace od rodilého mluvčího (viz 4.4), ale její zpřesnění by mohlo sloužit ke zlepšení kvality umělých dat. Tyto typy segmentace mohou být založeny na hledání lokálních extrémů trajektorií pohybu, rychlosti nebo zrychlení. Jak již bylo zmíněno v 3.3, nezanedbatelnou vlastností používaných dat je šum. Pro výsledný vzhled animace nemá šum s malou amplitudou velký vliv. v případě analýzy zaměřené na detekci lokálních extrémů a inflexních bodů v trajektorii však může tento šum úlohu ztížit (viz Obrázek 25).

Aby bylo možné s daty rychlosti a zrychlení lépe pracovat, přišel Mcdonald [19] s myšlenkou filtrace vysokých frekvencí pomocí dolní propusti. Rozdělil frekvenční spektrum pohybových dat do 3 sekcí: nízké frekvence, které obsahují většinu informace nesené pohybem do cca 2 Hz, vysoké frekvence nad 12 Hz, které obsahují pohyby nezpůsobené člověkem (např. elektromagnetický šum) a střední frekvence, které mohou být způsobeny lidským působením, ale nenesou významnou lingvistickou informaci. Cílem této části práce byl pokus replikovat rozdělení frekvenčního spektra použitých dat do obdobných sekcí. Experiment využívá data v angulární reprezentaci. Pohyb jednotlivých částí těla tak nezávisí na pohybu předchozích.

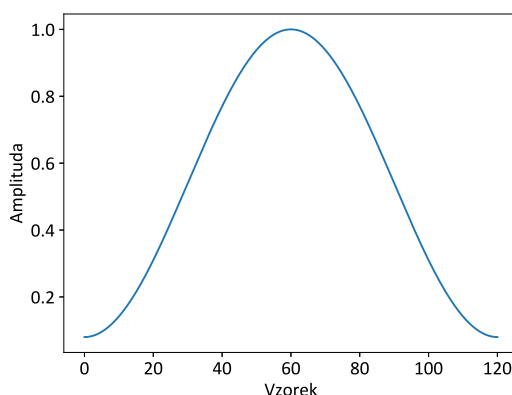


Obrázek 25: Poloha, rychlost a zrychlení pro náhodně vybranou část dat o délce 5 znaků

Frekvenční spektrum bylo získáno pomocí algoritmu rychlé Fourierovy transformace (FFT)<sup>16</sup>. Pro konečné signály se ve frekvenčním spektru projeví nežádoucí vliv obdélníkového okénka. Hlavní a postranní laloky znehodnotí informaci s nízkou amplitudou ve svém okolí. Tento jev lze potlačit použitím tzv. vyhlazovacího okénka. v praxi se používá větší množství oken, která se od sebe liší svojí frekvenční charakte-

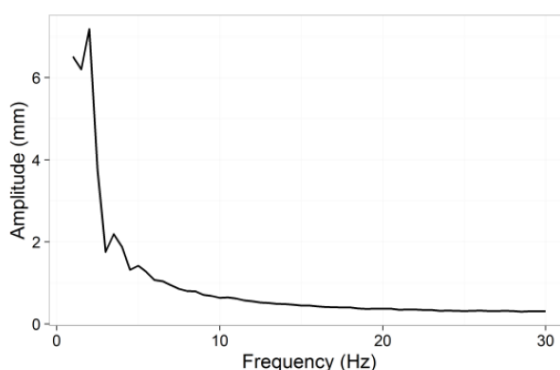
<sup>16</sup>Pro výpočet byla použita knihovna NumPy v1.16.3 FFT (pro Python3.7)

ristikou a proto jsou vhodné každé pro jinou úlohu. Okénko sice potlačí nežádoucí vliv, neodstraní jej však úplně. Na základě jednoduchých testů bylo pro tato data vybráno Hammingovo okénko.



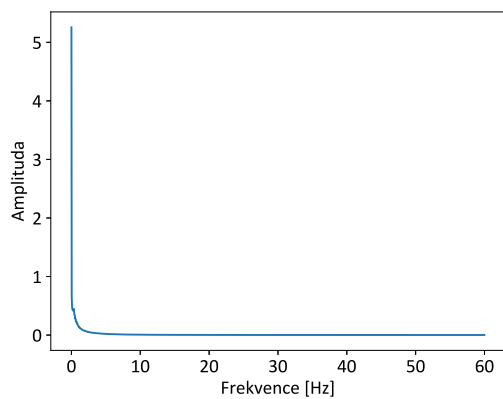
Obrázek 26: Tvar Hammingovo okénka

Na Obrázku 28 je znázorněn průměr frekvenčních spekter všech promluv a markerů. Amplitudy s rostoucí frekvencí strmě klesají bez známek prvků, které by mohly sloužit k rozlišení spektra do sekcí. Aby bylo možné ze spektra určit jednotlivé sekce, bylo by zapotřebí aby obsahovalo kromě šumu významné lokální extrémy. Na Obrázku 27 je vyobrazeno frekvenční spektrum, které analyzoval Medonald[19].

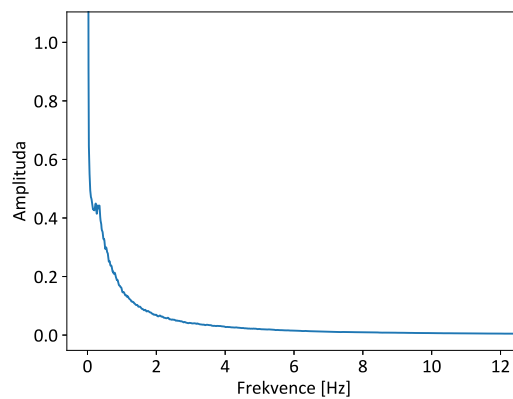


Obrázek 27: Frekvenční spektrum pro zápěstí, které využívá [19]

Myšlenkou, jak získat spektrum, ze kterého by bylo možné získat více informací, bylo využít pouze data ze statických částí promluv, které nenesou informaci. Cílem této metody bylo získat data s frekvenčním spektrem o malé amplitudě na nízkých frekvencích. Pokud by tomu tak bylo, mohly by být zpozorovány hledané prvky. Za nejstatičtější data jsou považovány klidové polohy mezi znaky ve slovníkových datech o délce do 70 snímků. Pro takto krátké záznamy jsou laloky okénka natolik výrazné, že překryjí okolní frekvence a ze spektra nelze z pohledu řešeného problému nic vyčíst, jak lze vidět na Obrázku 29. Při použití delšího úseku již přestává být póza statická a spektrum odpovídá dynamickým datům popsáno výše. Tato metoda tedy pro není účinná pro získání prvků potřebných k rozdělení spektra do zmíněných sekcí.

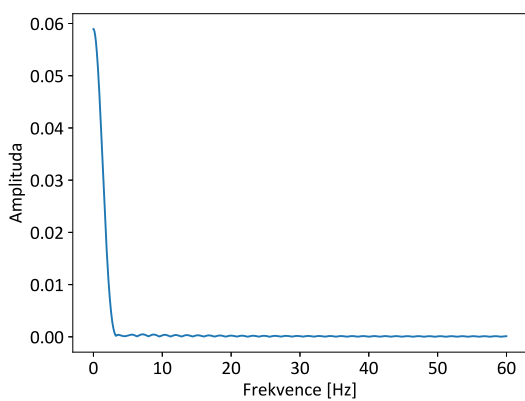


(a) Celé spektrum

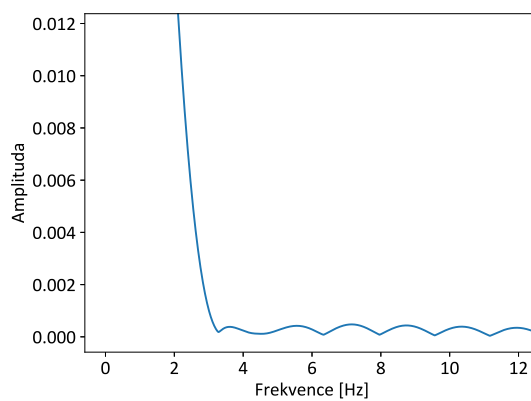


(b) Detail zkoumané části

Obrázek 28: Průměr frekvenčních spekter všech promluv a markerů



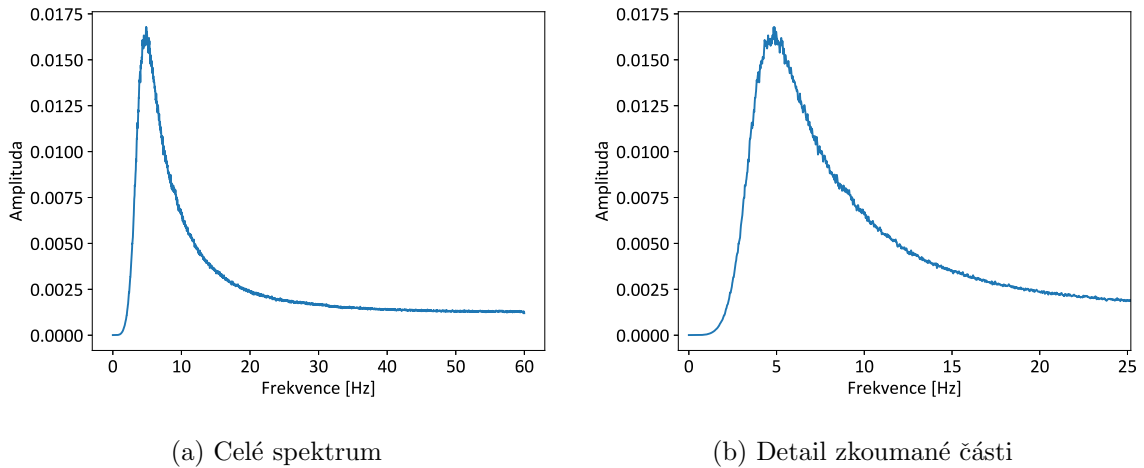
(a) Celé spektrum



(b) Detail zkoumané části

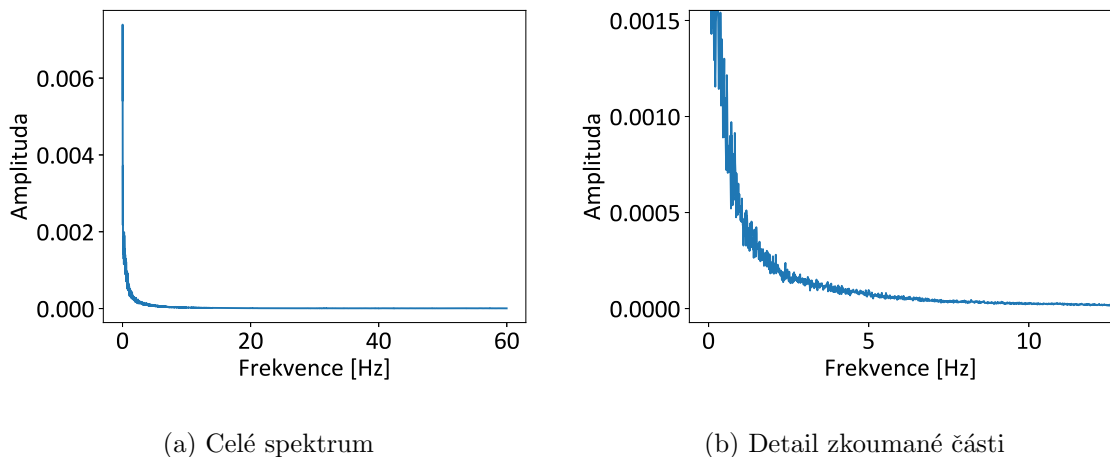
Obrázek 29: Průměr frekvenčních spekter všech markerů pro jednu statickou pózu

Dalším způsobem, kterým lze potlačit nízké frekvence, je použití filtru typu horní propusti. Pro tuto úlohu byl použit Butterworthův filtr<sup>17</sup>. Postupně byla upravována mezní frekvence v rozmezí 1-15 Hz a rychlost spádu v rozmezí 20-160 dB/dek. Příklad výsledného spektra je na Obrázku 30. Výsledná spektra pro jiná nastavení filtru mají obdobný průběh. Žádné z nich nejeví známky hledaných prvků.



Obrázek 30: Průměr frekvenčních spekter všech promluv filtrovaných horní propustí s mezní frekvencí 4 Hz a sklonem 60 dB/dek pro všechny markery

Šířka laloků frekvenčního spektra okénkové funkce závisí na šířce tohoto okénka, tedy na délce signálu. Data jednotlivých promluv lze sloučit do jedné dlouhé promluvy za cenu nespojitých přechodů. Použitá data (plynulá i slovníková) mají dohromady délku přibližně 313600 snímků o průměru jedné promluvy 5600 snímků. Šířka laloků pro takto dlouhá data pak bude zanedbatelná. Na Obrázku 31 je zobrazeno frekvenční spektrum pro trajektorii sloučených promluv. Spektrum již není ovlivněno laloky okénkové funkce, stále však nejeví známky hledaných prvků.



Obrázek 31: Průměr frekvenčních spekter všech markerů pro sloučené promluvy

<sup>17</sup>Pro výpočet byla použita knihovna SciPy v1.3.1 signal (pro Python3.7)

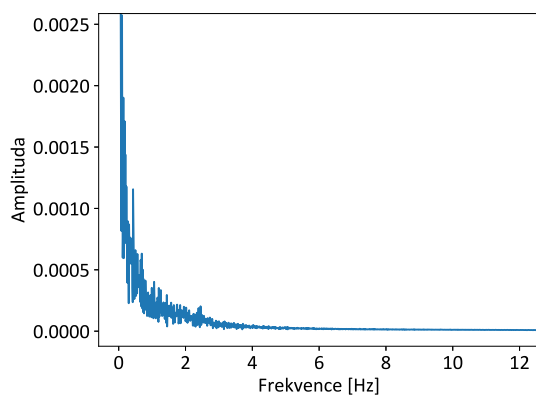
Doposud bylo spektrum počítáno jako průměr pro všechny klouby najednou. McDonald [19] uvádí hraniční hodnoty mezi nízkofrekvenční a střední sekcí pro vybrané části modelu kostry, který využívá (viz Tabulka 9). Na Obrázku 32 jsou vyobrazeny frekvenční spektra vybraných kloubů pro sloučené promluvy. z obrázku lze vidět, že spektra kloubů nesoucích více informace (např. zápěstí, loket) obsahují vyšší frekvence než spektra kloubů s menším obsahem nesené informace (např. kyčle, hlava), což odpovídá předpokladu [19]. Ve spektrech pro jednotlivé klouby lze v šumu pozorovat významnější lokální extrém. v žádném z nich však nelze jednoznačně určit hledané hranice hranice sekcí. v případě, že by v datech byly vyfiltrovány vyšší frekvence s hranicemi uvedených v Tabulce 9, ztratil by pohyb významnou část frekvencí a pohyb by byl nerealistický.

Výsledek žádné z použitých metod neodpovídá výsledkům které dosáhl McDonald [19] využitím spektra zobrazeného na Obrázku 27. Pro data používaná pro tuto práci se tedy tento experiment nepovedl zreplikovat. To je pravděpodobně způsobeno přílišnou závislostí metod na použitých datech nebo volbou vyhlazovacího okénka.

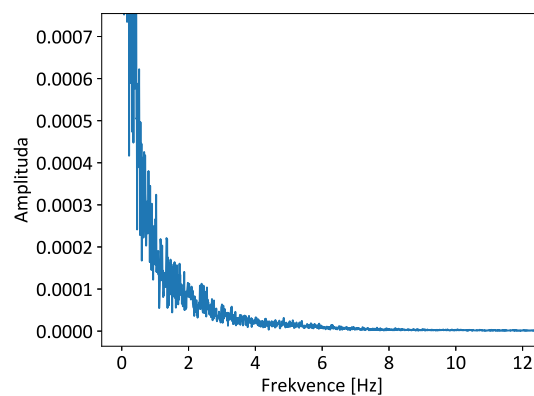
Kloub	Frekvence [Hz]
Boky	0,25
Pas	0,5
Horní část páteře	0,5
Krk	1,0
Ramena	1,0
Loket	2,0
Zápěstí	2,0

Tabulka 9: Hraniční hodnoty nízkofrekvenční sekce pro vybrané markery podle [19]

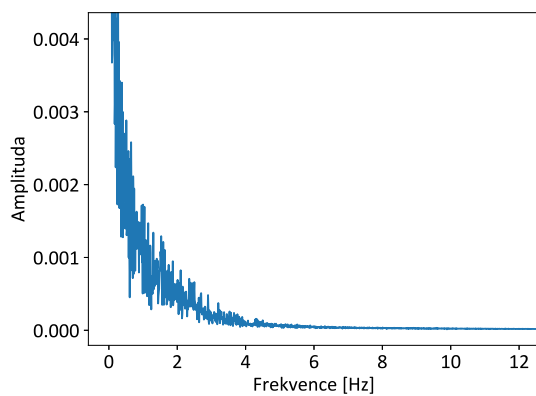




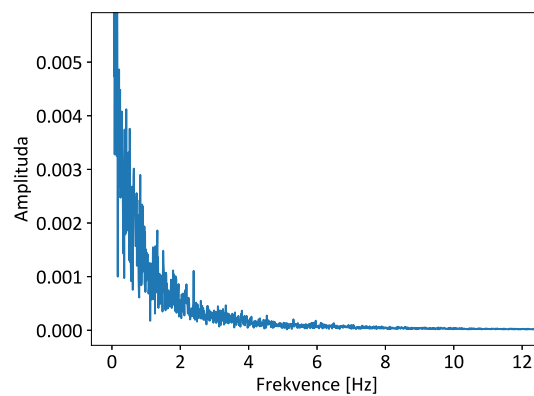
(a) Kyčle



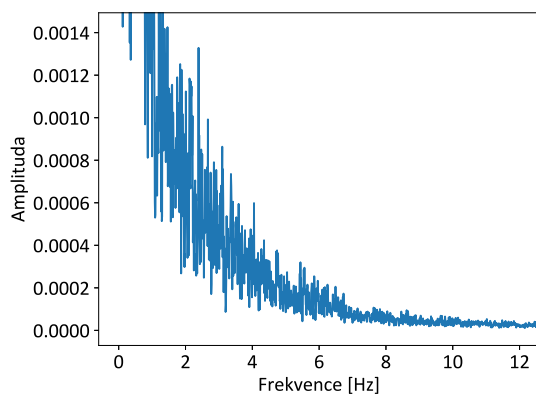
(b) Páteř



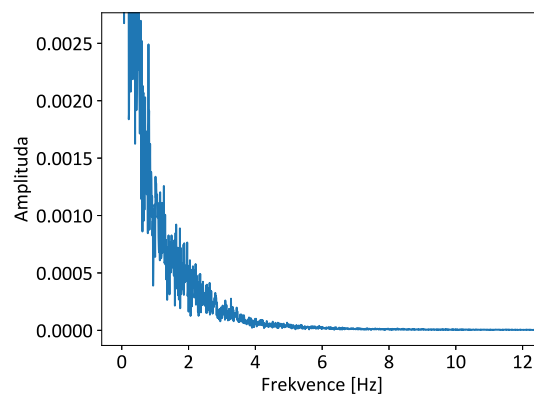
(c) Ramena (průměr)



(d) Lokty (průměr)



(e) Zápěstí (průměr)



(f) Hlava

Obrázek 32: Frekvenční spektra pro vybrané markery pro sloučené promluvy (Detail)

## 6 Závěr

Znakový jazyk je mateřským jazykem Neslyšících. Je pro ně tedy přirozenější komunikace ve znakovém jazyce nežli ve psané formě mluveného jazyka. Protože ne vždy je při tlumočení předem známa promluva a proces simultánního tlumočení není triviální, je proto snaha vyvinout automatickou syntézu znakového jazyka, která Neslyšícím umožní přístup k většímu množství informací v přirozené formě. Tvorba zvoleného typu syntézy vyžaduje velké množství dat. Těchto dat je však velký nedostatek. Nová data lze uměle pořídit pomocí procesu augmentace. Cílem této práce bylo analyzovat data, vybrat metody a měřitelné veličiny, které by mohly přispět k tvorbě augmentovaných dat či zhodnocení jejich přirozenosti.

Základními veličinami, které lze pro použitá data měřit, jsou rychlost a zrychlení. Lidské tělo se může pohybovat pouze s omezenou maximální rychlostí a zrychlením. Jedná se tedy o kritérium, které může umělá data vyhodnotit jako nerealistická. Vektor rychlosti je dále možné využít k vypočtení míry dynamiky, která může sloužit jako validace umělých dat jednotlivých částí těla. Podílem dynamiky levé a pravé ruky lze získat poměr, pomocí kterého lze pro promluvy spolehlivě určit, která ruka je mluvčím používaná jako dominantní. Dále byl tento poměr testován jako míra pro určení jednorukosti-dvourukosti jednotlivých znaků pro různé metody uspořádání dat. Tato metoda dosahuje až 99% přesnosti pro slovníková data, pro plynulé promluvy však díky vlivům okolních znaků nedosahuje přesnosti více než 88 %.

Strojově není možné přesně vyhodnotit přirozenost, ale lze však určit podobnost mezi dvěma znaky. Pokud je umělý znak dostatečně podobný reálnému, lze ho považovat za přirozený. v práci byly testovány míry podobnosti: euklidovská vzdálenost, dynamic time warping a korelace. Žádná z těchto metod nedokázala jednoznačně odlišit znak z vybrané množiny pro celé tělo. Při experimentu, kde byly měřeny pouze pohyby prstů uspěly všechny z těchto metod, nejlépe však metoda korelace. Dále byla pomocí korelace pravé a levé paže automaticky určována symetrie pohybu. Tato metoda se ukázala být více účinná pro data v angulárních souřadnicích než pro data v kartézské soustavě souřadnic.

Dále byl proveden pokus na základě [19] rozdělit frekvenční spektrum použitých dat do sekcí, které by mohly přispět k filtraci šumu a bylo by tak možné snáze detekovat extrémní trajektorie rychlosti a zrychlení. Tento experiment se však nepodařil replikovat se stejnými výsledky. Rozdíl může být způsoben přílišnou specifičností dat nebo volbou použitého vyhlazovacího okénka.

Stanovené cíle analýzy použitých dat byly splněny. Výše popsané metody mohou být použity jako jeden z validačních prvků uměle vytvořených dat. v budoucnu by bylo možné pokračovat v analýze informační hodnoty, kterou nesou jednotlivé části těla. To by mohlo být využito k heuristickému vážení těchto částí a tedy ke zlepšení výsledků výše popsaných metod. Dále je třeba v budoucnu analyzovat pohyby nemanuální složky znakové řeči. Jedná se o živou problematiku která se stále probírá a lze dále rozvíjet.

## Reference

- [1] J. Berke, "Self-identification in the deaf community," 2010. Dostupné online: <http://deafness.about.com/cs/culturefeatures1/a/bigdorsmalld.htm> (poslední přístup: 4. 7. 2020).
- [2] A. Macurová, "Poznáváme český znakový jazyk i.," *Speciální pedagogika*, vol. 11, no. 2, pp. 69–75, 2001. Dostupné online: <http://ruce.cz/clanky/8-poznavame-cesky-znakovy-jazyk-i> (poslední přístup: 4. 7. 2020).
- [3] autor neznámý, "Historie znakového jazyka ve světě," 2006. Dostupné online: <http://ruce.cz/clanky/59-historie-znakoveho-jazyka-ve-svete> (poslední přístup: 4. 7. 2020).
- [4] A. Macurová, "Proč a jak zapisovat znaky českého znakového jazyka," *Speciální pedagogika*, vol. 6, no. 1, pp. 5–19, 1996. Dostupné online: <http://ruce.cz/clanky/15-proc-a-jak-zapisovat-znaky-ceskeho-znakoveho-jazyka> (poslední přístup: 4. 7. 2020).
- [5] M. Holubová, "Prostor v českém znakovém jazyce," 2006. Dostupné online: <http://ruce.cz/clanky/42-prostor-v-ceskem-znakovem-jazyce> (poslední přístup: 4. 7. 2020).
- [6] J. Motejzíková, "Poznáváme český znakový jazyk v. - specifické znaky," *Speciální pedagogika*, vol. 13, no. 3, pp. 218–226, 2003. Dostupné online: <http://ruce.cz/clanky/189-poznavame-cesky-znakovy-jazyk-v-specificke-znaky> (poslední přístup: 4. 7. 2020).
- [7] P. Vysuček, "Poznáváme český znakový jazyk vi. - specifické znaky," *Speciální pedagogika*, vol. 14, no. 1, pp. 16–27, 2004. Dostupné online: <http://ruce.cz/clanky/190-poznavame-cesky-znakovy-jazyk-vi-specificke-znaky> (poslední přístup: 4. 7. 2020).
- [8] T. Nekvapil, J. a Sherman, "Foreigner talk," in *CzechEncy - Nový encyklopedický slovník češtiny*, 2007. Dostupné online: <https://www.czechency.org/slovník/FOREIGNERTALK> (poslední přístup: 4. 7. 2020).
- [9] V. Sutton, *Lessons in sign writing*. SignWriting, 1995.
- [10] T. Hanke, "Hamnosys-representing sign language data in language resources and language processing contexts," in *LREC*, vol. 4, pp. 1–6, 2004.
- [11] autor neznámý, "Horse art - eadweard muybridge, leland stanford and the hobby horse." Dostupné online: <https://www.your-guide-to-gifts-for-horse-lovers.com/muybridge.html> (poslední přístup: 4. 7. 2020).

- [12] D. R. Cho, “Brilliance (1985).” Dostupné online:  
<https://www.youtube.com/watch?v=M1cqc2AykmI>  
(poslední přístup: 4. 7. 2020).
- [13] G. Lucas, “Star wars: Episode i - the phantom menace,” 1999.
- [14] P. Jackson, “The lord of the rings: The two towers,” 2002.
- [15] J. Cameron, “Avatar,” 2009.
- [16] A. Szczesna, P. Skurowski, E. Lach, P. Pruszowski, D. Peszor, M. Paszkuta, J. Słupik, K. Lebek, M. Janiak, A. Polanski, and K. Wojciechowski, “Inertial motion capture costume design study,” *Sensors*, vol. 17, p. 612, 03 2017.
- [17] Česká televize, “Zprávy v českém znakovém jazyce.”  
<https://www.ceskatelevize.cz/porady/1097176961-zpravy-v-ceskem-znakovem-jazyce/>  
(poslední přístup: 4. 7. 2020).
- [18] M.-E.-F. Benchiheub, B. Berret, and A. Braffort, “Collecting and analysing a motion-capture corpus of french sign language,” 05 2016.
- [19] J. Mcdonald, R. Wolfe, R. Wilbur, R. Moncrief, E. Malaia, S. Fujimoto, S. Baowidan, and J. Stec, “A new tool to facilitate prosodic analysis of motion capture data and a data- driven technique for the improvement of avatar motion,” 05 2016.