5th Workshop on Spoken Language Technology for Under-resourced Languages, SLTU 2016, 9-12 May 2016, Yogyakarta, Indonesia

Automatic Technologies for Processing Spoken Sign Languages

Alexey Karpov [a]*, Irina Kipyatkova[a], Milos Zelezny[b]

[a]St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS), Russian Federation
[b]University of West Bohemia (UWB), Pilsen, Czech Republic

Abstract

Sign languages are known as a natural means for verbal communication of the deaf and hard of hearing people. There is no universal sign language, and almost each country has its own national sign language and fingerspelling alphabet. Sign languages use visual-kinetic clues for human-to-human communication combining hand gestures with lips articulation and facial mimics. They also possess a special grammar that is quite different from that of speech-based spoken languages. Sign languages are spoken (silently) by a hundred million deaf people all over the world and the most popular are American (ASL), Chinese, Brazilian, Russian, and British Sign Languages; there are almost 140 such languages according to the Ethnologue. They do not have a natural written form, and there is a huge lack of electronic resources for them, in particular, vocabularies, audio-visual databases, automatic recognition and synthesis systems, etc. Thus, sign languages may be considered as non-written under-resourced spoken languages. In this paper, we present a computer system for text-to-sign language synthesis for the Russian and Czech Sign Languages.

1. Introduction

At present, sign languages (SLs) are well known as a natural means for verbal communication of the deaf, hard-of-hearing people, and people who have some speaking difficulties. There is no universal sign language, and almost

---

* Corresponding author. Tel.: +7-812-328-0421; fax: +7-812-328-7081.
  E-mail address: karpov@iias.spb.su

each country has its own national sign language and fingerspelling alphabet. All the sign languages use visual-kinetic clues for human-to-human communication combining manual gestures with lips articulation and facial mimics. They also possess a specific and simplified grammar that is quite different from that of acoustic-based spoken languages.

Sign languages are spoken (silently) by a hundred million deaf people all over the world. In total, there are at least 138 living sign languages according to the Ethnologue catalogue, and many of them are national (state) or official languages of human communication in some countries like the USA, Finland, the Czech Republic, France, the Russian Federation (since 2013), etc. According to the statistics of medical organizations, about 0.1% of the population of any country is absolutely deaf and the most of such people communicate only by sign languages; many people, who were born deaf, even are not able to read. Additionally to conversational sign languages there are also fingerspelling alphabets, which are used to spell words (names, rare words, unknown signs, etc.) letter-by-letter. A fingerspelling system directly depends on the national alphabet of a country; also, there are both one-handed fingerspelling alphabets (e.g., in Russia, France, USA) and two-handed ones (e.g., in the Czech Republic, the UK, Turkey).

The most popular SLs with approximate numbers of native signers/speakers, who use a sign language as the first language of communication, according to the Ethnologue catalogue (http://www.ethnologue.com/subgroups/deaf-sign-language-0) and information from Wikipedia (https://en.wikipedia.org/wiki/List_of_sign_languages_by_number_of_native_signers), are:

- Chinese SL – about 20M signers (there is no exact information in catalogues, it is an estimation)
- Brazilian SL – about 3M signers;
- Indo-Pakistani SL – at least 2.7M signers (up to 10M according to some Internet sources);
- American SL (ASL) – 500K signers, mainly in the USA;
- Hungarian SL – 350K signers;
- Kenyan SL – 340K signers;
- Japanese SL – 320K signers;
- Ecuadorian SL – 188K signers;
- Norwegian-Malagasy SL – 185K signers;
- British SL (BSL) – 125K signers;
- Russian SL – 121K signers in the Russian Federation and up to 100K in other countries;
- French SL – 100K signers in France, plus about 50K in other French-speaking countries, etc.

Russian SL (RSL) is a native language for the deaf in Russia, Belarus, Ukraine, Kazakhstan, Moldova, also partly in Bulgaria, Latvia, Estonia, and Lithuania; there also exist some regiolects of Russian SL (RSL differs even in Moscow and St. Petersburg regions)[1]. Czech SL (CSL) is used by almost 10K deaf people in Czech Republic.

Family relationships of SLs are not well established because of the lack of linguistic research, but it is known that French SL family joins over 50 different SL in Europe (including Russian and Czech SLs), North and Latin America, and Africa; British and Arabic SL families include from 4 to 10 different SLs each.

All the sign languages do not have a natural written form, and there is a huge lack of electronic resources for them, in particular, vocabularies, multimedia and video databases, automatic recognition and synthesis systems, etc. Thus, sign languages may be considered as non-written under-resourced spoken languages. However, in the second half of the 20th century, some useful ways to represent sign languages in a written form have been proposed, which are called sign notation systems, for instance:

- Stokoe notation proposed by W. Stokoe[2]
- SignWriting developed by V. Sutton (http://www.signwriting.org)
- Hamburg Notation System (HamNoSys)[3]
- ASL-phabet developed by S. Supalla (http://www.aslphabet.com)
- Movement-Hold (M-H) notation proposed by S. Liddell and R. Johnson[4]
- Dimskis notation proposed by L. Dimskis[5]
- Si5s writing system proposed by R. Arnold (http://www.si5s.org)