Towards Russian Sign Language Synthesizer: Lexical Level

Alexey A. Karpov[1] and Miloš Železný[2]

[1] St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS), Russia
[2] The University of West Bohemia (UWB), Pilsen, Czech Republic
E-mail: karpov@iias.spb.su, zelezny@kky.zcu.cz

Abstract

In this paper, we present a survey of existing Russian sign language electronic and printed resources and dictionaries. The problem of differences in dialects of Russian sign language used in various local communities of Russia and some other CIS countries is discussed in the paper. Also the first version of a computer system for synthesis of elements of Russian sign language (signed Russian and fingerspelling) is presented in the given paper. It is a universal multi-modal synthesizer both for Russian spoken language and signed Russian that is based on a model of animated 3D signing avatar. The proposed system inputs data in the text form and converts them into the audio-visual modality, synchronizing visual manual gestures and articulation with audio speech signal. Generated audio-visual signed Russian speech and spoken language is a fusion of dynamic gestures shown by the avatar's both hands, lip movements articulating words and auditory speech, so the multimodal output is available both for the deaf and hearing-able people.

Keywords: Russian sign language, Russian signed speech, fingerspelling, sign language resources, computer synthesizer

1. Introduction

At present, sign languages are national languages of human-to-human communication in the USA, Finland, Czech Republic, etc. In the Russian Federation, official status of the sign language is lower and it is not a national communication means yet.

One of the major problems connected with the Russian sign language (RSL), which is regularly used by 1.5-2 million deaf people both in Russia and in some CIS countries (Ukraine, Belarus, Kazakhstan, etc.), consists in geographical vast of the country and existence of various dialects of RSL, for example, in Moscow, St. Petersburg, Novosibirsk, Vladivostok, Minsk, etc. Differences in dialects originate from a history of earliest schools for the deaf: the first one was organized in Pavlovsk near St. Petersburg in early XIX century (it still works) with teachers from France, and the first school in Moscow was opened much later and teachers were from Germany.

It is clear that differences in various versions of the sign language cause high ambiguities for translation and interpretation. An expert analysis shows that less than half of signs for the same words are similar in different dialects of RSL, these are simple manual gestures showing real objects in the field of vision (for example, "I", "YOU", "HEAD", "NOSE", etc.), however, most of abstract concepts highly depend on practice and traditions of local communities.

2. Russian Sign Language Resources

There exist several multimedia electronic dictionaries of RSL (some of them are available on-line and the rest are distributed by media):

1) "Thematic dictionary of the Russian sign language" developed by the Moscow organization of All-Russian society of the deaf in 2006 (www.deafmos.ru/info.phtml?c=24&id=1059), it has above 3K commonly used gestures on 4 DVDs, several human demonstrators are presented (Figure 1a).
2) "Russian Sign Language Explanatory Dictionary RuSLED" has over 2.5K gestures with etymology of the signs (Voskresenskiy, Gulenko & Khakhalin, 2009) and contains video data recorded in 2002 by the Inter-regional Rehabilitation Center for deaf people in Pavlovsk (Figure 1b).
3) Electronic learning system "Russian sign language. Basic course" (http://istina.inion.ru/NIOT/rgy.htm) created by the "Truth Center" in Moscow in 2001, it has up to 2K gestures (Figure 1c).
4) Interactive on-line dictionary DigitGestus (www.digitgestus.com) was collected in the Novosibirsk region in the late 90s.
5) New database recorded by the Novosibirsk State Technical University in 2010-2011 with above 3.2K signs (Grif et al., 2011a). It has a good quality of video data and one demonstrator (Figure 1d).
6) On-line RSL dictionary of the European project "Spreadthesign" (www.spreadthesign.com/ru) for 15 different sign languages of the world, it includes over 4.5K gestures of RSL (Figure 1e).
7) On-line RSL dictionary collected by the Stanford University (www.stanford.edu/dept/lc/rsl/). It has some hundreds of signs recorded in the USA.
8) On-line dictionary "Surdoserver" (http://surdoserver.ru), it has the same video data as in (3) plus an additional dictionary on Information Technology topic, it also has a version for mobile devices.
9) On-line dictionary "Surdoportal" (http://vorb.ru/ps/) has a small-sized dictionary for several technical topics.

Among the printed illustrated dictionaries of RSL, the following ones should be mentioned: books authored by Geilman (1979; 2001) from St. Petersburg (see Figure 1f), written by Zaitseva (2000), Fradkina (2001) from Moscow and by Dimskis (2002) from Minsk, etc.

(a) (b) (c) (d) (e) (f)

Figure 1: Variety of signs for the word “ONE” (digit) in different dictionaries of RSL

These electronic and printed dictionaries represent lexicon of RSL and illustrate differences in RSL dialects. An expert analysis of these electronic and printed dictionaries shows that only 30-40% gestures are similar in the Moscow and Petersburg dialects, though in some cases gestures can be understood, the same situation in other regions as well. Even in the Moscow region there are some RSL sub-versions in local communities and dictionaries are not unified and normalized, for example, the sign shown in Figure 1c (Figure presents various realizations of signs for the word “ONE”) taken from the electronic system “Russian sign language. Basic course”, created under support of the All-Russian society of the deaf in Moscow, differs from the sign for this word from other dictionaries originated from the Moscow region. So, it is not possible to say about one normalized RSL and unified automatic system for RSL analysis and synthesis.

Grammatical structure of RSL is not sufficiently studied and formalized yet to say on fully automatic text-to-sign language translation. Last years, some studies on structure of RSL grammar are made by linguists in the Moscow State University (Kibrik & Prozorova, 2007; Prozorova, 2009), in the Novosibirsk State Technical University (Grif & Demyanenko, 2011b), as well as by some other researchers working inside and outside of Russia (Voskresenskiy, Gulenko & Khakhalin, 2009; Kimmelman, 2009a; Kimmelman, 2009b; Mjasoedova & Filippovich, 2010). Unfortunately, the current scientific level and essential differences in semantic-syntactic structure of written/spoken and sign languages do not allow to perform machine translation from Russian texts to Russian sign language and there are no any models for automatic translation yet. In order to create such a model, it is required to use “deep” semantic, pragmatic and situational analysis and parsing of written phrases, however, at present there exist only superficial semantic analyzers because of imperfection of algorithms, databases and ontology for Russian and general complexity in grammatical and morphological structure of the language.

Our main goal is to develop a computer sign language synthesis system for the St. Petersburg version of RSL. At the given stage of the research we have developed a model for computer synthesis of signed Russian and fingerspelling, where input text processing is much easier.

3. Computer Synthesizer

There is quite large community aiming at computer processing of sign languages in Europe organized around several European projects. However, there is a lack of computer systems for RSL processing including sign language analysis and synthesis. Some models for signed Russian synthesis have been developed in Novosibirsk (www.vesti.ru/doc.html?id=358385&p=35&cid=1) and in Minsk (http://ont.by/news/our_news/0062482). Both are based on compilation of video fragments for whole phrases from pre-recorded video databases with one human demonstrator. However, an essential disadvantage of such systems is that produced video stream cannot show continuous signing and it is only isolated signs synthesis, because all the lexicon items are video fragments, where a demonstrator shows each sign independently from other signs and every gesture starts and ends with a neutral position of the hands (usually both hands are in an initial position below the belt). In the case of real continuous signing people do not go through this initial position after each gesture, but only in the beginning and the end of the whole phrase. So, video dictionaries cannot serve as the basis for construction of continuous sign language and signed speech synthesizers for human-computer interaction.

Therefore, animated characters or virtual humans (avatars) are more adequate for this task. A database of animated 3D gestures allows compiling phrases in a sign language in the continuous manner keeping smooth transitions from sign to sign.

Since 2009, SPIIRAS Institute and the University of West Bohemia have been developing a multimodal text-to-sign language system for RSL. It is originally based on a 3D signing avatar for signed Czech speech and language (Krňoul et al., 2008), parameters of which and hand movements are controlled by the codes of Hamburg Notation System (Prillwitz et al., 1989; Hanke, 2004). The proposed system inputs data in the text form

and converts them into the audio-visual modality, combining visual manual gestures and articulation with audio-visual speech, so the multimodal output is available both for deaf and hearing-able people. The main software components of the multimodal system are:

1) Text processing module that inputs and processes text phases and generates word labels, phonemic and visemic transcriptions, as well as a stream of inner codes of hands movements from the dictionary.
2) Text-to-speech system for spoken Russian that generates auditory speech signal corresponding to the text (Hoffmann et al., 2007).
3) Virtual 3D model of human's head with controlled lips articulation, mimics and facial expressions (Zelezny et al., 2006).
4) Bimodal audio-visual "talking head" that integrates the speech synthesizer and the virtual head model and synchronizes lip movements with synthesized auditory speech signal taking into account natural asynchrony between audio and visual speech modalities (Karpov et al., 2009).
5) Virtual 3D model of human's upper body (Kanis & Krňoul, 2008) with both hands, movements of which are controlled by the codes of HamNoSys notation.
6) Multimodal computer synthesizer that synchronizes and integrates all the components for automatic generation of auditory speech, visual speech (articulation) and manual gestures of signed speech and language.

Figure 2 shows general architecture of the multimodal system, its main components and interaction between them. In the proposed system, synchronization of audio-visual speech with manual signs is controlled using time stamps of start and end of spoken words generated by the auditory speech synthesizer. Since natural speech has a higher tempo than the corresponding manual gestures, then the signing avatar speaks and articulates isolated spoken words and the system waits for the following acoustic word until completion of the current gesticulation (if no any sign for a word in the system's dictionary it is spelled as a sequence of finger sings by the avatar's right hand). By this way continuous gesticulation of the whole phase is provided.

The sign language synthesizer based on high-quality virtual 3D avatars has a lot of merits:

- It allows a user to see generated visual data from different sides and viewing angles that leads to better understanding of spatial information, for example distance between the hands and the body or hands each from other.
- It is possible to add new items into the dictionary quite easily; it is an animated virtual human, so there is no requirement to record one human demonstrator in the same dress, haircut and make-up with similar lighting conditions and equipment.
- It can produce a continuous stream of visual gestures without transitions through a neutral position of hands and there are no seen borders between adjacent signs.
- It is possible to change one virtual avatar to another one and to create new models of human beings.
- It is able to show synthesized signed phrases on a screen with any required speed, slowing down or speeding up the visual stream.

The multimodal system is aimed not only for deaf, deaf-mute and hearing impaired people, but is useful for hearing people as well. It is a universal multimodal computer system for synthesis both Russian spoken language (audio-visual modality) and the sign language (visual modality). Generated audio-visual signed Russian speech and language is a fusion of dynamic gestures shown by the avatar's both hands (or only by the right hand in the case of Russian fingerspelling), lip movements articulating spoken words and acoustic speech. Many deaf people are able to read speech by lips and to understand phrases even without manual gestures. Acoustic spoken language is a natural speech modality for communication with hearing-able people. Avatar's lips articulation synchronized with audio stream helps to improve both intelligibility and naturalness of generated speech.

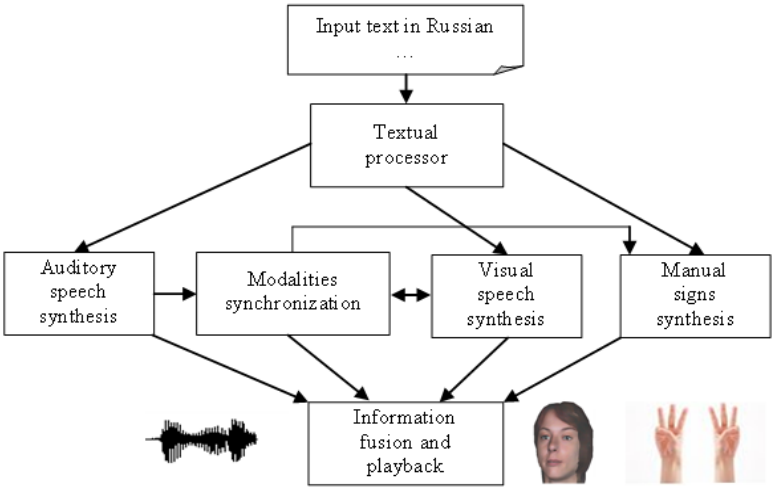

Figure 2: Architecture of the multimodal synthesizer for audio-visual spoken language and signed Russian

4. Conclusions

We presented the survey of RSL resources and the first version of the text-to-sign synthesizer for signed Russian and fingerspelling. It is the universal multimodal system for synthesis of Russian spoken language (audio-visual modality) and signed Russian (visual modality) aimed both for the deaf and hearing-able people.

Demonstrations of the multimodal synthesis system for Russian fingerspelling and elements of RSL are available on-line: www.spiiras.nw.ru/speech/demo/daktilrus.avi and www.spiiras.nw.ru/speech/demo/signlang.avi.

Qualitative user evaluation of the system was made in the end of 2011 with the help of some representatives of the All-Russian society of the deaf in St. Petersburg. They said on novelty and urgency of the system and positively estimated intelligibility and naturalness of lips articulation of the talking head and recognizability of manual gestures of the virtual avatar. At the same time they expressed some requirements for the future work, i.e. to use lexical items of RSL from the books and dictionaries (i.e., prepared by Gejlman) created in St. Petersburg region only and to enforce further research on text-to-sign language machine translation.

5. Acknowledgements

We are very grateful to Dr. Zdeněk Krňoul from the University of West Bohemia in Pilsen for provision of the 3D signing avatar and its parameters adjustment, as well as to Olga Novoselova, who is the head of the All-Russian society of the deaf in St. Petersburg, and her colleagues for assistance in evaluation of the synthesizer and suggestions on system's improvement.

This research is supported by the Ministry of Education and Science of Russia in the framework of the Federal Target Programme «R&D in Priority Fields of S&T Complex of Russia for 2007-2013» (contract No. 11.519.11.4025), by the grant of the President of Russia (project No. MK-1880.2012.8), and by the Ministry of Education of the Czech Republic, project No. ME08106.

6. References

Dimskis, L.S. (2002). *Learning sign language* (Изучаем жестовый язык), Moscow: Academia, 128 p.

Fradkina, R.N. (2001). *Speaking hands: Thematic dictionary of sign language of the deaf of Russia* (Говорящие руки: Тематический словарь жестового языка глухих России), Moscow, 598 p.

Gejlman, I.F. (1979). *Specific means of communication of the deaf. Fingerspelling and mimics* (Специфические средства общения глухих. Дактилология и мимика). 4 volumes. Leningrad.

Gejlman, I.F. (2001). *Get to know: Manual speech* (Знакомьтесь: Ручная речь), Moscow: Zagrej, 172 p.

Grif, M.G., Demyanenko, Y.A., Korolkova, O.O., Tsoy, Y.B. (2011a). Development of Computer Sign Language Translation Technology for Deaf People. In *Proc. IFOST-2011*, Harbin, China, pp. 674--677.

Grif, M.G., Demyanenko, Y.A. (2011b). Russian text analysis in a system of text to Russian sign language translation (Анализ русского текста в системе компьютерного сурдоперевода на русский разговорный жестовый язык). *NSTU Proceedings*. Novosibirsk, Russia, 63 (1), pp. 83--90.

Hanke, T. (2004). HamNoSys - representing sign language data in language resources and language processing contexts. In *Proc. LREC-2004*, Lisbon, Portugal, pp. 1--6.

Hoffmann, R., Jokisch, O., Lobanov, B., Tsirulnik, L., Shpilewsky, E., Piurkowska, B., Ronzhin, A., Karpov, A. (2007). Slavonic TTS and SST Conversion for Let's Fly Dialogue System. In *Proc. SPECOM-2007*, Moscow, Russia, pp. 729--733.

Kanis, J., Krňoul, Z. (2008). Interactive HamNoSys Notation Editor for Signed Speech Annotation. In *Proc. LREC-2008*, Morocco, pp. 88--93

Karpov, A., Tsirulnik, L., Krnoul, Z., Ronzhin, A., Lobanov, B., Zelezny, M. (2009). Audio-Visual Speech Asynchrony Modeling in a Talking Head. In *Proc. Interspeech-2009*, Brighton, UK, pp. 2911--2914

Kibrik, A.A., Prozorova E.V. (2007). Referencial choice in Russian sign language (Референциальный выбор в русском жестовом языке). In *Proc. Dialog-2007*, Moscow, Russia, pp. 220--230.

Kimmelman, V. (2009a). Parts of speech in Russian Sign Language: The role of iconicity and economy. *Sign Language & Linguistics*, 12(2). pp. 161--186.

Kimmelman, V. (2009b). Reflexive pronouns in Russian Sign Language and Sign Language of the Netherlands // MA thesis, Universiteit van Amsterdam.

Krňoul, Z., Kanis, J., Železný, M., Müller, L. Czech Text-to-Sign Speech Synthesizer (2008). In *Proc. MLMI Workshop*, Brno, Czech Republic, Springer LNCS 4892, pp. 180--191

Mjasoedova, M.A., Filippovich, Yu.N. (2010). A system for automated translation of signed speech (Система автоматизированного перевода жестовой речи). In *Proc. CLAIM Workshop*, Moscow: MSTU, Russia.

Prillwitz S., et al. (1989). *HamNoSys. Hamburg Notation System for Sign Languages. An introductory guide.* (International Studies on Sign Language and Communication of the Deaf; Hamburg: Signum, 46 p.

Prozorova, E.V. (2007). Russian sign language as the object of linguistic research (Российский жестовый язык как предмет лингвистического исследования). *Voprosy yazykoznaniya*, 1, pp. 44--61.

Voskresenskiy, A., Gulenko, I., Khakhalin, G. (2009). RuSLED Dictionary as Tool for Semantic Study // *Computational Linguistics and Intellectual Technologies*, Moscow, Russia, pp. 64--69.

Zajtseva, G.L. (2000). *Signed speech. Fingerspelling* (Жестовая речь. Дактилология), Moscow: Vlados, 192 p.

Zelezny, M., Krnoul, Z., Cisar, P., Matousek, J. (2006). Design, implementation and evaluation of the Czech realistic audio-visual speech synthesis // *Signal Processing,* 86(12), pp. 3657—3673.