Publishable summary

Abstract

Deaf communities revolve around sign languages as they are their natural means of communication. Although deaf, hard of hearing and hearing signers can communicate without problems amongst themselves, there is a serious challenge for the deaf community in trying to integrate into educational, social and work environments. The overall goal of SignSpeak is to develop a new vision-based technology for recognizing and translating continuous sign language to text, being this the first step to approach this technology at levels already obtained in similar technologies such as automatic speech recognition or statistical machine translation of spoken languages. New knowledge about the nature of sign language structure from the perspective of machine recognition of continuous sign language will allow a subsequent breakthrough in the development of a new vision-based technology for continuous sign language recognition and translation. Existing and new publicly available corpora will be used to evaluate the research progress throughout the whole project. Project partners and other public information can be found on the project website: www.signspeak.eu; contact: info@signspeak.eu.

Figure 1. SignSpeak Logo

SignSpeak Specifications

1. **Multimodal system**. Due to the many simultaneous ‘channels’ of signed languages (two hands, face, head and upper body), the system will extract information not only from the dominant hand, but also from the non-dominant hand and from the facial expression and body position (shoulders, elbows and chest).
2. **More natural.** The signer will speak without wearing gloves or other types of sensors or markers. The entire process will be vision based: non-invasive system.
3. **Robustness and self-adaptation to the changing ambient conditions**. During the project, research will be targeted at the development developing robust feature extraction techniques: the hands are signing often in front of the face (occlusions), and standard face detection methods often fail due to strong facial expressions, head tilt and head turns: that is a challenging task in sign language recognition. Additional research will be carried out to allow the system to work independently of the background colour and the signers’ clothes and brightness.
4. **Signer-independency**. Thanks to the statistical approach for gesture and sign language recognition, the system will be gender and age-independent similar to robust automatic speech recognition systems. Signer independence also implies **pronunciation, language modelling adaptation and the usage of speaker adaptation techniques**. In addition, the clothing are not going to be controlled (just avoid white clothing).
5. **Contextual translation.** The system will carry out continuous sign language translation within a context, not merely identifying isolated signs.
6. **Multilingual.** One scientifically challenging task is that there are many different sign languages in Europe, with only a few described grammars. The suggested recognition and translation systems will be based on statistical methods for modelling the appearance and the grammar: these methods have proven to be the most powerful techniques for automatic speech recognition and machine translation in the last years. In addition, the advantages of using these data driven methods gives the technology robustness and scalability to other languages by using different training data. Therefore, although the project will be developed

to work with NGT, the system will be also trained and tested to smaller extent in German Sign Language (DGS).

7. **Spatial Reference Handling.** A challenging task will be to analyse the spatial information containing the entities created during the sign language discourse; it could reduce the ambiguity of words that are typically a problem in translation systems (e.g. pronouns). References in signing space occur quite often to refer to previously deposited objects in the virtual signing space.
8. **Software Integration.** The different prototypes developed separately for multimodal visual analysis, sign language recognition and translation will be integrated by communicating the different applications under a common framework. A graphical user interface will be designed and developed for the easy use of the system.
9. **Context-domain of the translations**. For the Sign Language of the Netherlands, SignSpeak works with video records (Corpus-NGT) created by posing 15 questions to 46 pairs of signers, accounting around 90 hours of clips; these questions elicit 'discussions' about issues related to the deaf community and deafness. After analysing the observations (word-frequency) in the Corpus NGT, it has been selected this 'discussion' domain for targeting the SignSpeak translations. Regarding German Sign Language (DGS), a smaller corpus is built up by recording the weather forecast in a German TV-station; therefore, the context domain is going to be the weather forecast.
10. **Real time factor around 20 for translating NGT**. The final system is not going to run in real time: a real time factor of 20 means that 6 seconds of video records will take 2 minutes for providing the translation. An online demonstration is foreseen for translating the sign language of The Netherlands (NGT), in contrast to the other focused sign language (DGS), where the demonstration will be done by offline evaluations due to the smaller size of the Corpora available for training the system.
11. **Vocabulary size** (for NGT) around 4.000 words.

Progress at the end of the first year of the project

A conceptual scheme of the work planned is presented in next figure.

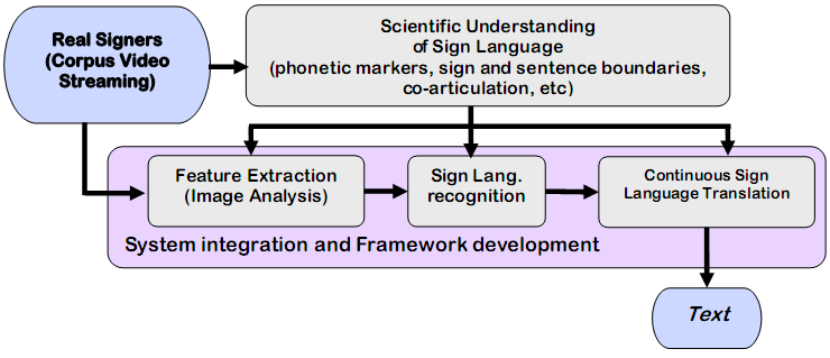

Figure 2. Conceptual scheme of the work planned in SignSpeak project

The following points summarise the progress made in the different blocks:

- **Video Corpus:** The features of the existing Corpora for setting up the specifications of the system have been studied, as well as for defining the requirements of gloss, translations,

sentence boundaries and non-manual annotations. Considerable effort has been invested in carrying out all these annotations, which will also take place during the second year of the project.

- **Scientific understanding of signed languages**: Literature study revealed that there are no 'hard' cues thus far for sentence boundaries to be exploited for sign recognition research, and a new approach is required whereby combinations of cues are analyzed. Several of our linguistic studies have shown that some lexical items are predictive for sentence boundaries, given that these items often occur at the start or the end of sentences, although not consistently. Thus far, lexical items and prosodic cues were analyzed separately in past studies. Both have shown to be predictive, however, not sufficiently to be able to detect sentence boundaries. Prosodic cues and lexical cues should therefore be combined to predict sentence boundaries for automatic sign language translations. The literature study revealed that video analyses and linguistic analyses can be mutually informative to gain further insight in the exact phonetic/prosodic cues present at sentence boundaries and should be exploited in further studies.
- **Feature extraction**: a Baseline Prototype for the multimodal visual analysis has been developed integrating hand and face tracking. The hand tracking method intrinsically allows a quantitative characterization of hand shapes; this avenue will be further pursued during the coming months. The face tracking method allows the quantification of certain aspects of facial expressions such as 3D head orientation and eye and mouth apertures. In addition, spatiotemporal features have been extracted by different approaches and will be integrated in next Advanced Prototype of the multimodal visual analysis to be delivered at the end of the second year of the project.

Figure 3. Feature extraction and expression quantification

- **Sign Language Recognition**: a Baseline Prototype for sign language recognition has been developed, which allows the recognition of isolated and continuous sign language data. It offers many configuration possibilities and will allow for the recognition of different signed languages in the future, such as the Corpus-NGT database (Sign Language of the Netherlands or NGT) or the RWTH-PHOENIX-v2.0 database (German Sign Language: GSL). The current prototype has been trained on appearance-based image features only, without any tracking features. The next steps will include the integration of more sophisticated features from the baseline prototype for multimodal visual analysis introduced before.
- **Sign Language Translation to text**: a Baseline Prototype for sign language translation has been developed by employing statistical sign language translation system based on a state-of-the-art hierarchical decoder. It works on continuous sign language, allowing for gaps in the translation by working on a CFG grammar structure. We also tested syntactically motivated methods in pre- and post-processing using additional monolingual data by means of a morpho-syntactic analyser (Morphisto) and a deep syntactic parser (Stanford parser) for German.

With regard of dissemination activities, partners involved in SignSpeak project co-organised along with Dicta-Sign partners (another EC funded project working in sign language recognition), two dissemination workshops at the most prestigious conferences in the domain of Language Resources and Computer Vision:

- CSLT 2010: SignSpeak partners RWTH and CRIC are organizing committee member of the "Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies (CSLT 2010)", Valletta, Malta, May 22nd-23rd, 2010. Organized as a Language Resources and Evaluation Conference (LREC 2010) post-conference workshop.

  URL: http://www.sign-lang.uni-hamburg.de/lrec2010/cfp.html

- SGA 2010: SignSpeak partners RWTH and ULg are organizing committee member of the "International Workshop on Sign, Gesture, and Activity (SGA 2010)", Hersonissos, Heraklion, Crete, Greece, Sep 11th, 2010. Organized as a European Conference on Computer Vision (ECCV 2010) satellite workshop.

  URL: http://personal.ee.surrey.ac.uk/Personal/R.Bowden/SGA2010/

In addition, papers from the SignSpeak consortium have been accepted at LREC 2010, CSLT 2010, BMVC 2009 and ESSV 2009.

Potential impact and use: Towards a Sign-Language-to-Spoken-Language Translation System

The interpersonal communication problem between signer and hearing community could be resolved by building up a new communication bridge integrating components for sign-, speech-, and text-processing. To build a sign-to-speech translator for a new language, a six component-engine must be integrated as shown in next figure, where each component is in principle language independent, but requires language dependent parameters/models. The models are usually automatically trained but require large annotated corpora. In SignSpeak, a theoretical study will be carried out about how the new communication bridge between deaf and hearing people could be built up by analyzing and adapting the ASLR and MT components technologies for sign language processing.

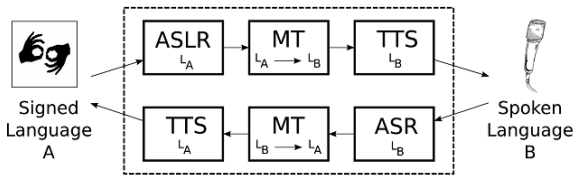

Figure 4. Complete six components-engine necessary to build a Sign-To-Speech system: automatic sign language recognition (ASLR), automatic speech recognition (ASR), machine translation (MT), and text-to-speech/sign (TTS).

Once the different modules are integrated within a common communication platform, the communication could be handled over 3G phones, media center TVs, or video telephone devices. The following application scenarios would be possible:

- e-learning of sign language;
- automatic transcription of video e-mails, video documents, or video-SMS;
- video subtitling and annotation.