DICTA-SIGN - Sign Language Recognition, Generation, and Modelling: A Research Effort with Applications in Deaf Communication

Eleni Efthimiou[1], Stavroula-Evita Fotinea[1], Thomas Hanke[2], John Glauert[3], Richard Bowden[4], Annelies Braffort[5], Christophe Collet[6], Petros Maragos[7], and François Goudenove[8]

[1]Institute for Language and Speech Processing, [2]Universität Hamburg, [3]University of East Anglia, [4]University of Surrey, [5]LIMSI/CNRS, [6]Université Paul Sabatier, [7]National Technical University of Athens, [8]WebSourd

E-mail: [1]{eleni_e,evita}@ilsp.gr, [2]thomas.hanke@sign-lang.uni-hamburg.de, [3]J.Glauert@uea.ac.uk, [4]R.Bowden@surrey.ac.uk, [5]annelies.braffort@limsi.fr, [6]collet@irit.fr, [7]maragos@cs.ntua.gr, [8]francois.goudenove@websourd.org

Abstract

Here we present the components and objectives of Dicta-Sign, a three-year FP7 ICT project that aims to improve the state of web-based communication for Deaf people by allowing the use of sign language in various human-computer interaction scenarios. The project researches and develops recognition and synthesis engines for sign languages at a level of detail necessary for recognising and generating authentic signing. In this context, Dicta-Sign aims at developing several technologies demonstrated via a sign language-aware Web 2.0, combining work from the fields of sign language recognition, sign language animation via avatars, sign language linguistics, and machine translation, with the goal to allow Deaf users to make, edit, and review avatar-based sign language contributions online, similar to the way people nowadays make text-based contributions on the Web.

Dicta-Sign supports four European sign languages: Greek. British, German, and French Sign Language and differs from previous work in that it aims to integrate tightly recognition, animation, and machine translation. All these components are informed by appropriate linguistic models from the ground up, including lexical and grammar modelling, manual and non-manual features.

1. Rationale

The development of Web 2.0 technologies has made the WWW a place where people constantly interact with another, by posting information (e.g. blogs, discussion forums), modifying and enhancing other people's contributions (e.g. Wikipedia), and sharing information (e.g., Facebook, social news sites). Today's predominant human-computer interface, is relatively manageable for most Deaf people: The use of a language foreign to them is restricted to single words or short phrases. The graphical user interface, however, puts rather severe limitations on the complexity of the human-computer communication, and therefore it is expected that it will be replaced in many contexts by human language interaction. Obviously, a far better command of the interface language is required here than in graphical environments. Most Deaf people would therefore be excluded from this future form of human-computer communication unless the computer is able to communicate in sign language. Moreover, exclusion is already experienced with regard to interpersonal communication between Deaf individuals, given the current lack of translation tools to support SL-to-SL but also oral-to-SL and SL-to-oral applications. Sign language videos are not a viable alternative to text, for two reasons: Firstly, they are not anonymous – individuals making contributions can be recognized from the video and therefore limits those willing to contribute. Secondly, people cannot easily edit and add to a video that someone else has produced, so a Wikipedia-like web site in sign language is currently not possible. In order to make the Web 2.0 fully accessible to Deaf people, sign language contributions must be displayed by an animated avatar, which addresses both anonymisation and easy editing.

2. The Dicta-Sign project

Dicta-Sign is a project aimed at developing the technologies required for making sign language-based Web contributions possible, by providing an integrated framework for sign language recognition, animation, and language modelling. It targets four different European sign languages: British (BSL), German (DGS), Greek (GSL) and French (LSF), and develops three proof-of-concept prototypes: a search-by-example sign language dictionary, a sign language-to-sign language translator, and a sign language-based Wiki.

A key aspect of the Dicta-Sign project is the creation of parallel corpora with detailed annotations in the four above-mentioned signed languages. These not only greatly aid the development of language models for both recognition and animation, but also allow for the direct spatio-temporal alignment of equivalent utterances across the four languages, which is useful for creating machine translation algorithms in a sign language-to-sign language translator.

3. Objectives

One of the main objectives of Dicta-Sign is to develop an integrated framework that allows contributions in the four sign languages of the project. Users make their contributions via webcams. These are recognized by the sign language recognition component and converted into a linguistically informed internal representation, which is used to animate the contribution with an avatar, and to translate it into the other respective three sign languages.

Other objectives include the development of the world's first parallel multi-lingual corpus of annotated sign language data; the development of advanced sign language annotation tools that integrate recognition, translation, and animation; the provision of large cross-lingual sign language dictionaries; and the advancement of the state of the art in computer vision and sign language recognition, sign language generation, sign language linguistic modelling and sign language translation.

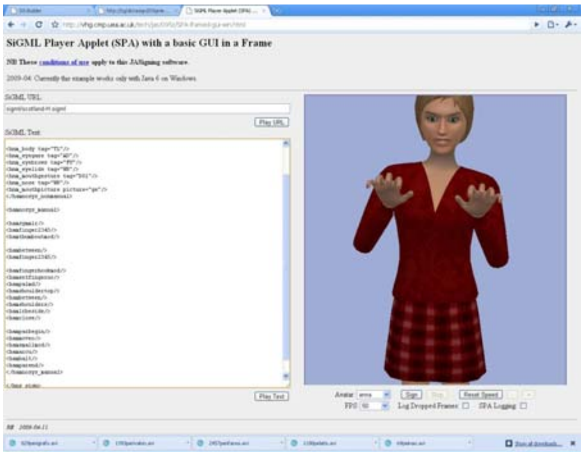

Figure 1: Avatar frontal view demonstrating incorporation of non-manuals in the sign synthesis engine: eyes, eyebrows, mouth and body posture participate in sign articulation

Dicta-Sign is working closely with the Deaf communities in the countries of the project partners throughout the lifecycle of the project to ensure that its goals are met and to evaluate user acceptance.

4. Research Domains

Research activities within Dicta-Sign expand from sign language recognition to sign synthesis and animation, linguistic modelling and development of annotation tools.

4.1 Sign Language Recognition

Despite intensive research efforts, the current state of the art in sign language recognition leaves much to be desired. Problems include a lack of robustness, particularly when low-resolution webcams are used, and difficulties with incorporating results from linguistic research into recognition systems. Moreover, because signed languages exhibit inherently parallel phenomena, the fusion of information from multiple modalities, such as the hands and the face, is of paramount importance. To date, however, relatively little research exists on this problem (Ong & Ranganath, 2005). The features that serve as input to the recognition system comprise a mix of measurements obtained by statistical methods, and geometrical characterisations of the signer's body parts. In order to make the feature extraction process robust even when the image comes from commodity webcams, the computer vision algorithms need to operate on multiple scales. Moreover, the basic feature extraction processes need to be combined with statistical and learning-based methods, such as active appearance models for facial expression tracking (Cootes et al., 2001; Papandreou & Maragos, 2007).
Sign language is inherently multimodal: both hands move in parallel, while the face and body exhibit grammatical and prosodic information (Neidle et al., 2000). Hence, sign language recognition must deal with the problem of fusing multiple channels of information.

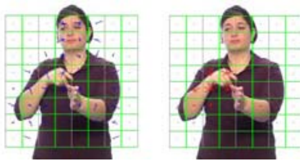

Figure 2: Figure showing HOG/HOF features (Histogram of Oriented Gradients, Histogram of Optical Flow) from a single frame of Sign Footage

Given the current state of the art in sign language recognition, one cannot expect the system to recognize the full range of expressiveness in signed languages. We deal with this limitation in two ways: First, the prototype application is domain-specific, with a restricted vocabulary of no more than 1500 signs. Second, the system employs a dictation-style interface (hence the name "Dicta-Sign"), where the user is presented with the closest-matching alternatives if a sign is not recognized reliably.
The output of the recognition component is converted into a linguistically informed representation that is used by the synthesis and language modelling components, respectively.

4.2 Synthesis and Animation

In the Dicta-Sign project, the internal representation of sign language phrases is realized via SiGML (Elliott et al., 2000), a Signing Gesture Markup Language to support sign language-based HCI, as well as sign generation. The SiGML notation allows sign language sequences to be defined in a form suitable for execution by a virtual human, or avatar, on a computer screen. The most important technical influence on the SiGML definition is HamNoSys, the Hamburg Notation System (Hanke, 2004), a well-established transcription system for sign languages. The SiGML notation incorporates the HamNoSys phonetic model, and hence SiGML can represent signing expressed in any sign language.
One of the most difficult problems in sign synthesis is converting a linguistic description of the signed utterance into a smooth animation via inverse kinematics, with proper positioning of the hands in contact with the body, and generating realistic prosodic features, such as appropriate visual stress. To this end, the Dicta-Sign corpus, does not only encompass phonetic and grammatical information, but also prosodic information. Together with the features derived from the visual tracking and recognition component, this allows for greatly increased realism in the animations.

4.3 Linguistic Modelling

Linguistic modelling will develop a coherent model from the phonetic up to the semantic level of language representation, envisaged to be language-independent in most aspects. Dicta-Sign aims to extend modelling capabilities toward a common representation of sign language grammar and the lexicon -or alternatively two coherent representations- to accommodate both sign language recognition and synthesis. Overall, this represents a major advance over previous work, since language modelling has been largely neglected particularly in the recognition field.

4.4 Annotation Tools

Although some tools exist for specifically processing signed languages, such as iLex, none of these tools currently provide any kind of automated tagging, so the annotation process is completely manual.

An experimental version of the AnCoLin annotation system allows some image processing tasks to be initiated from within the annotation environment and to compare the results with the original video (Braffort et al., 2004; Gianni et al., 2007). It also connects to a 3D model of the signing space, but still lacks a coherent integration into the annotation workflow. It is expected that one of the major outcomes of the Dicta-Sign project will be greatly improved annotation tools, with image processing and recognition integrated into the annotation workflow. Their long term utility can be judged by the uptake by other sign language researchers.

4.5 Sign Language Corpora

A substantial corpus is needed to drive automatic recognition and generation, so as to obtain sufficient data for training and language representation. The quality and availability of sign language corpora has improved greatly in the past few years (Efthimiou & Fotinea, 2007; Neidle & Sclaroff, 2002). Yet, to date, multi-lingual sign language research has been hampered by the lack of sufficiently large parallel sign language corpora. One of the most important goals of Dicta-Sign is to collect the world's first large parallel corpus across four signed languages (Greek, British, German, and French).

This corpus will be annotated, showcase best practices for sign language annotations, and be made available to the public.

Figure 3: Two handed sign articulation by neutral body posture

5. Expected Outcomes

Expected outcomes of the project expand to both prototype systems of SL technologies and SL resources, and include:

- A parallel multi-lingual corpus for four national sign languages – German, British, French and Greek (DGS, BSL, LSF and GSL respectively) – of a minimum of three hours signing in each language,
- A substantial dictionary of at least 1500 signs for each represented sign language,
- A continuous sign language recognition system that achieves significant improvement in terms of coverage and accuracy of sign recognition in comparison with current technology; furthermore this system will research the novel directions of multimodal sign fusion and signer adaptation,
- A language generation and synthesis component, covering in detail the role of manual, non-manual and placement within signing space,
- Annotation tools which incorporate these technologies providing access to the corpus and whose long term utility can be judged by the up-take by other sign language researchers,
- Three bidirectional integrated prototype systems which show the utility of the system components beyond the annotation tools application,
- A showcase demonstrator which exhibits how integration of the different components can support user communication needs.

6. Proof-of-concept Prototypes and Project Demonstrator

Three proof-of-concept prototypes will be implemented and evaluated within Dicta-Sign:

- A Search-by-Example system will integrate sign recognition for isolated signs with interfaces for searching an existing lexical database.
- An SL-to-SL translation prototype will pioneer a controlled-vocabulary sign language-to-sign language translation on the basis of the parallel language resources developed within the project.
- A Sign-Wiki will be developed providing the same service as a traditional Wiki but using sign language.

As a showcase of the different technologies developed within Dicta-Sign, an SL-to-SL terminology translator will be developed to serve as project demonstrator.

7. Conclusion

Today, still living in the atmosphere of the "European Year of Equal Opportunities for All," it is important that drastic measures are taken to prevent new barriers from arising, as new forms of communication establish their role in the society at large. Dicta-Sign will be a key technology to promote sign language communication, and to provide Web 2.0 services and other HCI technologies to Deaf sign language users, an important linguistic minority in Europe, so far excluded from these new developments.

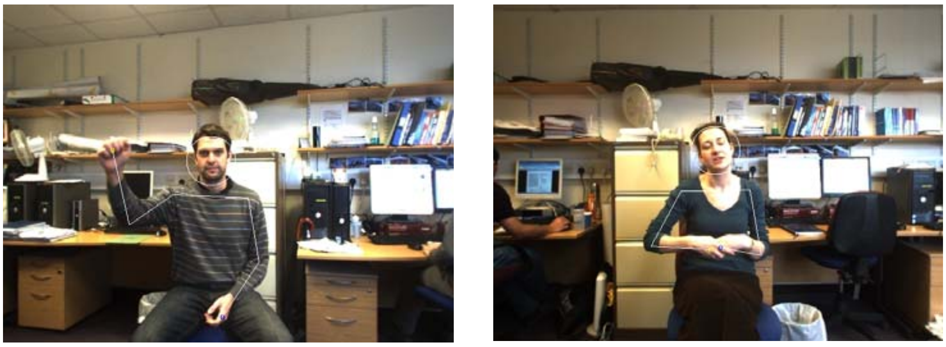

Figure 4: Examples of coarse pose estimation during Signing

Figure 5: Motion estimation and segmentation

As the field of sign language technology is still very young, it is beyond the scope of a three-year project to catch up completely with mainstream language technology, and to deliver end-user products. Nevertheless, Dicta-Sign is poised to advance significantly the enabling technologies by a multidisciplinary approach, and to come close enough to let designers of future natural language systems fully take sign languages into account.

8. Acknowledgements

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 231135.

9. References

Braffort, A., Choisier, A., Collet, C. (2004). Toward an annotation software for video of Sign Language, including image processing tools and signing space modelling. In *Proceedings LREC 2004*.

Cootes, T. F. and Edwards, G. J. and Taylor, C. J. (2001). Active Appearance Models. In *IEEE Trans. PAMI*, vol.23, no.6, pp. 681--685.

Efthimiou, E. & Fotinea, S-E. (2007). GSLC: Creation and Annotation of a Greek Sign Language Corpus for HCI. In *Proceedings of 12th International Conference on Human-Computer Interaction*: *Universal Access in HCI*, Part I, HCII 2007, LNCS 4554, pp. 657--666.

Elliott, R., Glauert, J.R.W., Kennaway, J.R., and Marshall, I. (2000). Development of Language Processing Support for the Visicast Project. In *Proceedings ASSETS 2000 4th International ACM SIGCAPH Conference on Assistive Technologies*, Washington DC, USA, 2000.

Gianni, F., Collet, C., Dalle, P. (2007). Robust tracking for processing of videos of communication's gestures. In *Proceedings International Workshop on Gesture in Human-Computer Interaction and Simulation (GW 2007)*, Lisbon, Portugal, May 2007.

Hanke, T. (2004). HamNoSys - representing sign language data in language resources and language processing contexts. In *Proceedings LREC 2004, Workshop proceedings: Representation and processing of sign languages*. Paris: ELRA, 2004, pp. 1--6.

Marshall, I., Sáfár, E. (2005). "Grammar Development for Sign Language Avatar-Based Synthesis", In *Proceedings HCII 2005, 11th International Conference on Human Computer Interaction* (CD-ROM), Las Vegas, USA, July 2005.

Neidle, C., Kegl, J., MacLaughlin, D., Bahan, B. and Lee, R.G.. (2000). The Syntax of American Sign Language: Functional Categories and Hierarchical Structure. Cambridge, MA. MIT Press.

Neidle, C. and Sclaroff, S. (2002). Data collected at the National Center for Sign Language and Gesture Resources, Boston University. Available online at http://www.bu.edu/asllrp/ncslgr.html

Ong, A.C.W. and Ranganath S. (2005). Automatic Sign Language Analysis: A Survey and the Future beyond Lexical Meaning. In *IEEE Trans. PAMI*, 27(6): pp. 873--891, 2005.

Papandreou, G. and Maragos, P. (2007). "Multigrid Geometric Active Contour Models. In *IEEE Trans. Image Processing*, vol.16(1), pp. 229--240.