FP7 - Information and Communication Technologies

Collaborative Project

Dicta-Sign

Sign Language Recognition, Generation and Modelling
with application in Deaf Communication

Grant Agreement Number 231135

Deliverable D8.1: Project demonstrator: Sign-Wiki

Leading Partner: ATHENA RC/ILSP-Institute for Language and Speech Processing

Version Final / February 2012
# Table of Contents

0. Foreword to D8.1 ............................................................................................................. 3
1. Rational for the DICTA-SIGN demonstrator ............................................................. 3
2. Demonstrator implementation issues............................................................................ 3
3. Demonstrator Graphical User Interface...................................................................... 4
   3.1 Basic Sign Wiki content page ........................................................................... 5
   3.2 Content Structure Panel ............................................................................... 6
   3.3 Signing Avatar Panel ................................................................................... 6
   3.4 Edit/Input Mode ......................................................................................... 6
4. User Interaction options................................................................................................ 7
   4.1 Content Structure Panel – Viewing Mode ....................................................... 7
   4.2 Edit/Input Mode ........................................................................................... 8
   4.3 Content Structure Panel – Edit/Input Mode .................................................... 9
   4.4 Sign Insertion and Editing............................................................................... 11
      4.4.1 Kinect Input/Recognition Module ......................................................... 11
      4.4.2 Dicta-Sign Lexicon Search .................................................................... 13
      4.4.3 Sign Builder ......................................................................................... 13
   4.5 Translation service .......................................................................................... 15
5. Concluding remarks .................................................................................................... 16
0. Foreword to D8.1

As already stated in the project DoW document, Deliverable D8.1 is intended to describe the project demonstrator, as one of the deliverables of WP8. Following the lines of the Dicta-Sign DoW, “in WP8 the systems developed in WP7, in the framework of project prototype integration work, will be fine-tuned to provide a demonstrable showcase”.

In what follows, we describe the Dicta-Sign demonstrator, which is based on the project prototype described in detail in deliverable D7.3: Sign-Wiki Prototype.

1. Rational for the DICTA-SIGN demonstrator

During the Dicta-Sign project lifecycle, It has been emphasised in many occasions that Dicta-Sign’s goal has been to develop the necessary technologies that make Web 2.0 interactions in sign language possible, raising the obstacles connected with the use of video in the Internet and mainly answering those questions connected with the lack of anonymity and the ease of editing of pieces of video. In the framework of the Dicta-Sign project, these barriers were faced by developing sign language recognition and animation technologies which exploit richly annotated parallel resources from four sign languages and experiment with generally applicable language models for sign language vocabulary and grammar, in order to allow the development of Web based applications capable to support communication in Sign Language.

All project achievements have been integrated in three proof-of-concept prototype systems, composing work in WP7.

The project prototypes have been i) a Search-by-Example Tool, described in deliverable D7.1, ii) a Sign Look-up Tool, described in deliverable D7.2, and iii) a Sign-Wiki prototype, described in D7.3.

During the project, Dicta-Sign has been opened to extensive user evaluation procedures as well as to independent Sign Language expert evaluation in order to benefit from both user and project external expert feedback. As a result, the Sign-Wiki application gathered most of the interest of the Deaf project evaluators and it was also recognised as the most innovative of all its foreseen prototypes from the part of external experts. This view was also shared by the year-2 project evaluators. These reactions in combination with the fact that the Sign-Wiki prototype inherently incorporates all technologies developed in the project and also makes use of the in-project created sign language resources, make it an ideal project demonstrator.

2. Demonstrator implementation issues

The Sign-Wiki demonstrator shares the prototype technical features reported in D7.3 and it is built using conventional Web technology, augmented with various special purpose SL processing modules. On the server side, the Sign-Wiki Web site is implemented using the Django framework, supported by a PostgreSQL database. The client-side of the Web site is implemented using HTML and Javascript, supported by the jQuery library, and several other small jQuery-related library modules. The Kinect Input/Recognition service implemented in the demonstrator is described in project deliverable D7.2.

The server-side database fulfills two functions:

• Providing access to the Dicta-Sign Lexicon for a set of over a thousand concepts, in each of the four SLs supported by the project.
• Acting as a dynamically developing repository, allowing for storage of current and previous content versions created by the users.

For specially designed XML formats employed in the Sign-Wiki demonstrator, reference is made in section 3 of D7.3.

To use the Sign-Wiki the user needs just a conventional personal computer system and a web browser (Windows or Mac OS X, with Firefox). To do Kinect-based SL input (described in greater detail later), a Microsoft Kinect™ input device is required, attached either to the client computer itself or to another local computer, physically located close the main client system, and connected via the local network to that system. The local computer hosting the Kinect device also needs to have the special Kinect-service software, developed within the project (at UniS), installed on it. Apart from this, the Sign-Wiki depends only on standard client-side Web software technology.

3. Demonstrator Graphical User Interface

The user initiates a session on the Sign-Wiki demonstrator by entering its URL: http://signwiki.cmp.uea.ac.uk/dictasign/, via a link, into a Web browser.

This URL takes the user to the demonstrator home page, shown in Figure 1, which displays a variety of links into the main body of the system, allowing the user to:

• Access currently available documents, i.e. pages.
• Access previous versions of those documents.
• Work with links for each version of a document.
• Work with documents previously prepared elsewhere by uploading them to the Sign-Wiki.
• Install the HamNoSys fonts required for editing or creating content.

Figure 1: Sign Wiki home page
3.1 Basic Sign Wiki content page

The user may access content in one of the existing pages by selecting the wished link, as shown in Figure 2.

The row of icons at the top of the page, constitutes the page toolbar, the exact contents of which may vary with context.

To sketch the functionalities, we provide below a description of each toolbar icon.

Home icon 

takes the user back to the Sign-Wiki home page.

The following group of three icons allow the user to control the level of detail in the Content Structure panel:

- Section — so that sentences and signs are hidden
- Section and Sentence — so that signs are hidden
- Full — with no hidden elements.

The next group of icons correspond to the four national sign languages represented in the Dicta-Sign parallel corpus:

- French Sign Language (LSF)
- Greek Sign Language (GSL)
- British Sign Language (BSL)
- German Sign Language (DGS)
The selected SL is indicated by a border on the corresponding national flag icon. These icons also lead to translation functionalities incorporated in the system (see next section).

The final, rightmost, group of toolbar icons vary depending on the current access mode for the page. When in Viewing mode, as in Figure 2, there is a single icon, allowing the user to request a switch to Edit/Input Mode .

Below the toolbar, the GUI consists of two panels always present during user interaction: the Virtual Human Signer (Signing Avatar) panel on the right, where the user may view signed content by means of a virtual signer, and the Content Structure panel on the left, where the user may view content (tree-)structure, and hence navigate through the content and control its presentation.

3.2 Content Structure Panel

The content of a Sign-Wiki page is built from sequences of individual signs. Signs may be grouped under "linguistic structure" nestable items whose arguments are in turn signs or linguistic structures. At the top level they come under "header" items which were introduced in view of generating automatic outlines for the wiki pages, but do not have any other linguistic purpose.

The user access modes (Viewing and Edit/Input mode, referred to in details in the next section) influence the GUI as follows: When the page is in Edit/Input mode, the right-hand boundary of the Content Structure panel is moved to the right, allowing a third central Edit/Input Management panel to be inserted between the two primary panels.

3.3 Signing Avatar Panel

The Signing Avatar panel is located on the right of the Sign-Wiki page. Its main component being the Java applet panel provided by the JASigning SL synthesis and performance package developed in WP3, it allows the user to view SL sequences for all or selected parts of the page content presented in the Content Structure panel on the left. A small set of controls include:

- Play – performs, or replays, the current selection.
- Stop – stops the avatar performance prematurely.
- Slower – reduces the speed of the avatar's performance.
- Faster – increases the speed of the avatar's performance.
- Reset – resets the speed of the avatar's performance to its standard value.

3.4 Edit/Input Mode

In the Edit/Input mode the View mode's Edit (Pencil ) toolbar is replaced with a View mode switching button (the Eye icon ), together with a set of icons which provide access to facilities available only in Edit/Input mode:

- Save (Disc icon ) – is used to commit updates to the page at any stage during editing.
• Lexicon Search (Binoculars icon) – is used to search the Dicta-Sign lexicon for a sign or signs to be added to the page content.

• Kinect Input (Kinect icon) – is used to initiate input signs directly from the user, via a Kinect tracking device.

The Sign-Wiki's Kinect input/recognition module is invoked using the Kinect Input button in the toolbar. This module architecture is essentially the same as that used in the Search-by-Example Tool and Sign Look-up Tool, described in detail in deliverables D7.1 and D7.2. This is a (non Web) client-server architecture, in which the main Kinect server is provided by a computer system which must:

• have a Kinect motion capture device attached;
• be physically adjacent to the primary Sign-Wiki client computer (so that a user can perform a sign sequence in front of the Kinect);
• have the Dicta-Sign Sign Lexicon recognition service software (developed in WP1 and WP2) installed on it; and
• have a LAN connection to the primary Sign-Wiki client.

The Kinect server computer system may actually be the same as that running the primary Sign-Wiki client, but it need not be, provided it is physically close enough to the user who must sign in front of it.

The L icon, when activated, allows the user to insert and specify linguistic structures to wrap around the signs listed, and thereby build a more grammatical tree structure for the sentence.

4. User Interaction options

4.1 Content Structure Panel – Viewing Mode

The Content Structure panel on the left of the Sign-Wiki page presents a structured view of the page content to the user allowing for navigation within the page and selection of units for presentation by the virtual human signer.

The global level of detail at which the page structure is displayed can be controlled using the appropriate toolbar buttons by exploiting the level of detail provided by the small disclosure (+/-) buttons that appear to the right of each node in the Structure display, apart from an individual sign node – that is, each node with its own internal structure (page, section, sentence, or other linguistic structure). Each individual sign is labeled with its spoken language gloss name.

To view the page content, or any structural unit within the page, using the virtual human signer, the user simply selects by clicking on the required node, so as the virtual human signer initiates presentation of the selected sequence sign by sign. As each individual sign is presented, its entry in the tree Structure on the left is highlighted in blue, thus providing visual feedback to the user, also allowing for verification of the organisation of the content.
4.2 Edit/Input Mode
For any given Sign-Wiki content page, a switch from Viewing mode to Edit/Input mode (Figure 3) is achieved by clicking the Pencil icon that appears at the page's toolbar (as shown in Figure 2).

In this mode the View mode's Edit (Pencil) toolbar is replaced with a View mode switching button (the Eye icon), together with several more icons, providing access to facilities available only in Edit/Input mode. Here the user may execute the following actions:

- Save updates to the page at any stage during editing.
- Search the lexical resources of the Dicta-Sign lexicon for a sign or signs to be added to the content of the page under editing.
- Interact with the system to directly input signs, via a Kinect tracking device.

The Lexicon Search and Kinect Input functions are described in later subsections below.
4.3 Content Structure Panel – Edit/Input Mode

The use of the Content Structure panel, on the left of the window, when the page is in Viewing mode was described above (§4.1). In this subsection we describe the additional capabilities that become available when the page is in Edit/Input mode, as shown in Figure 4.

In this mode the Content Structure panel gains the ability to act as a Drag-and-Drop based structure editor. The user can select any node within the content tree structure and move it to a new position in the tree by dragging it there. More precisely, such a move is permitted provided it would not violate the tree's structural integrity: for example, the user is prevented from moving a section into the middle of a sentence. In accordance with standard Drag-and-Drop conventions, if the Alt/Option key is held down while dragging then the source node is copied rather than moved to the target location.

In this mode, the Viewing mode capabilities are still available to the user. In particular, it is possible to view any unit within the structure as often as the user desires while editing proceeds.

As can be seen in Figure 4, the Switch into Edit/Input Mode causes some buttons to appear on the right of each node in the Content Structure panel. These provide access to additional editing functions:

- **Delete** (Cross button 🚫) – used to delete the content represented by the given node from the Sign-Wiki page (guarded by an “Are you sure?” modal dialog).

- **Insert** (Star button 🌟) – triggers a popup menu allowing a skeleton new content node to be inserted adjacent to the given node.

- **Edit** (Pencil button 🖋) – triggers a popup menu showing the available sign editing methods.

When the Insert (Star) button is used to insert a new skeleton node, the newly inserted node's content may then be defined using any of the supported content
definition techniques appropriate to a node of the given kind. In the case of a new structural node this is a matter of using the new skeleton node’s own Insert or Edit buttons to expand the structure further recursively to the required extent. In the case of a new sign node, it is a matter of using the new (empty) sign’s own Edit button to create its definition.

The Edit (Pencil) button—which is provided only for Sign nodes, not for structural nodes—may support a range of input methods, at the project demonstrator environment, as e.g. the Sign Builder, an interactive, graphically based, sign definition module, described below (§4.4.3).

In Figure 5 below, the example in Figure 3b is elaborated exploiting the linguistic structure of enumeration for LSF. Figure 5a depicts an example of natural signing and 5b shows how the enumeration structure is edited in the edit mode.

Figure 5: (a) Head movement in natural enumerations for LSF, (b) Head movement imposed for the enumeration linguistic structure in LSF
4.4 Sign Insertion and Editing
At the most basic level, the creation of new content in a SignWiki page is a matter of inserting new signs into the page structure. The previous subsection identified the ways in which sign insertion is triggered. This subsection describes the currently available sign, or sign sequence, insertion methods. The content insertion options available to the user involve:

- Kinect input by exploitation of Sign Recognition technologies
- Search in the Dicta-Sign Lexicon
- Sign Builder use
- Use of the HamNoSys symbols

Both the first and second of these methods allow identification of signs, if already included in the Dicta-Sign lexicon. The Sign Builder allows the user to create a sign definition by identifying its individual features (handshape, orientation, location, movement, etc.) interactively. As well as the creation of new signs, the Sign Builder also allows the user to edit signs already included in the page, using the same set of interactive techniques.

4.4.1 Kinect Input/Recognition Module
The Kinect Input/Recognition Module is based on the Kinect-based motion capture and recognition technology that has been developed in WP1 and WP2, as mentioned earlier (§3.4).

The Sign-Wiki's Kinect Input/Recognition module allows the user to perform a sign language sequence using signs from the Dicta-Sign Lexicon. It include both reading-style continuous as well as dictation-style recognition capabilities. To allow the system to segment this performance into individual signs the user must perform it in a form of “dictation” mode, that is, relatively deliberately and with a brief pause at the temporal boundary between one sign and the next. The input/recognition service produces for each sign in the input sequence a short list of possible matches against entries in the Dicta-Sign lexicon, each list being ranked so that the most likely match appears first.

When the Sign-Wiki user invokes the Kinect input/recognition module, and upon receipt of the recognition results (lists), the Kinect Input Sequence Review mode is initiated (Figure 4), in which the user is given the opportunity to review them and select between the possible matches for each sign. The sequence is presented to the user in a sentence structure comparable to that used in the Content Structure panel on the left of the page. But here the entry for each sign has, in addition to its gloss label in the appropriate national spoken language, a pair of left and right arrow buttons. These buttons can be used to move back and forth through the list of entries from the Dicta-Sign lexicon that are possible matches for the sign performed by the user. Each time the user moves a new sign in the list, that sign is performed by the virtual human signer, so that the user can decide which (if any) of the possible matches is the one intended.

At the end of this review process, or indeed at any intermediate stage in it, the user may insert some or all of the input sequence, as currently displayed, into the page by dragging the required node(s) from the central Review panel to the chosen location in the Content Structure panel on the left.

The buttons at the bottom of the Review panel provide the following functions:

- Preview – has the entire input sequence, as currently displayed, performed by the virtual human signer.
• Finish – terminates the Kinect Input/Recognition module, dismissing the central panel, returning the Sign-Wiki page to Edit/Input mode.

• Redo Kinect Input – requests a new input/ recognition cycle by the Kinect input/recognition server.

The last of the options may be invoked either because the user is dissatisfied with the results produced for the last input/recognition cycle and wishes to re-attempt it, or simply because it is desired to input a completely new sequence.
4.4.2 Dicta-Sign Lexicon Search
The Dicta-Sign Lexicon Search module is invoked by means of the Search (Binoculars) button in the toolbar. This causes the Lexicon Search panel to be displayed in the central region of the page, as shown in Figure 5. The user enters the search keyword (“house” in the example in Figure 5) into the text entry field at the top of the Search panel. The system then attempts to match this keyword against the Wordnet concept definitions for all the Dicta-Sign Lexicon entries, and the matching entries (if any) are displayed in the Search Results area of the central panel. The user may then insert any of these entries into the current page by dragging their thumbnail icons to the desired point in the Content Structure panel on the left of the page. When the Finish button is clicked the central Search panel is dismissed and the Sign-Wiki returns to Edit/Input mode.

4.4.3 Sign Builder
The SignBuilder is invoked from the Edit popup menu for an individual sign in the Content Structure Panel, as explained above (§2.6). This brings up the Sign Builder panel in the centre of the page, as shown in Figure 8. The panel provides a sequence of sub-panels allowing the user to specify separate phonetic-level features of the required sign – using diagrams borrowed from the HamNoSys tutorial, since the project has standardised on HamNoSys as the basis for its phonetic-level sign representation. Six phonetic features may be specified for the sign:

- Handshape
- Thumb modifier
- Finger bending
- Extended Base Finger Direction
- Palm Orientation
- Location

Figure 7: Dicta-Sign Lexicon Search
As can be seen in Figure 8, the Sign Builder panel contains a button for each of these features. Each button acts as a toggle for the display/hiding of the graphic showing the range of possible settings for the given feature.

Figure 9 shows the Handshape selector in use. A range of 12 possible handshapes is displayed. The selection mechanism provides a high level of feedback to the user: as the pointer moves over the entries in the selector the virtual human signer on the right is used to display whichever entry the pointer hovers over from moment to moment, allowing the user to see in real-time the result that will be given if he/she actually selects the given entry, by clicking on it. When the pointer moves away from the selection area the virtual human signer shows the currently selected feature.

A later stage in the sign building process is shown in Figure 10: by this stage a new Location (shoulder level, central) has already been selected and the Palm Orientation is now being selected. When the required features of the sign have been established to the user's satisfaction, the user can either accept it, that is, insert it at the currently selected location in the Content Structure (Save button) or discard it (Cancel button).

As integrated in the Sign-Wiki demonstrator, the Sign Builder can be used interactively, as has just been shown, to define a complete posture, although it does not yet include support for the definition of movements, and in fact it does not currently support the full range of postural detail that can be expressed in HamNoSys. The Sign Builder can also be used to edit one or more postural features of a previously established sign, both for one-handed signs and for two-handed signs with left-right symmetry – criteria which are met by over half of the entries in the Dicta-Sign Lexicon.
4.5 Translation service

The flag icons indicating the environment's sign languages in the basic Sign-Wiki content page toolbar GUI (3.1), also allow the user to control a partial translation facility. As has already been described, the individual signs in a Sign-Wiki page can be created, and edited, in a variety of ways. One common source for these signs is the Dicta-Sign parallel Lexicon, containing over 1000 signs in each of the four sign languages. If the user indicates a wish to switch from one sign language – by clicking one of the three currently unselected SL icons – then those signs in the current page
which have been drawn from the Dicta-Sign lexicon, will be “translated” into the newly selected SL, achieved by substituting for each one the appropriate entry from the parallel lexicon.

It should be emphasised that this feature provides a translation facility that is no more than partial, in the following respects:

- It applies only to signs drawn from the Dicta-Sign lexicon, although the Sign-Wiki provides other means to create signs.
- Even in the case of a sign drawn from the lexicon, the translation is necessarily impossible, if the user has modified the definition of that particular instance (a capability that the Sign-Wiki does provide).

Initially, the page is in View mode, allowing the user to view its contents, as already described previously.

In Figure 11 the partial translation of the GSL sentences depicted in Figure 3 is shown for DGS.

5. Concluding remarks

Deliverable D8.1 presents the Dicta-Sign demonstrator, focusing on the end user’s approach to the project proof-of-concept prototypes. Under this light, the main aim here is to underline the range of actions and interaction possibilities that are finally offered to signing users of Web 2.0, resulting from research work in the framework of the DICTA-SIGN project.

Innovation in respect to communication via sign language is summarized in the following:

I) **Sign-Wiki users may view SL information** uploaded by other individuals. This may involve information in one’s own sign language or may require translation support in order to be comprehended.

II) In the latter case, the user may find support by a **Sign-Look-up translation module**, which allows search of signs in four sign languages. Multilingual correspondences of the same concept increase the possibility of its understanding.

III) **The user may edit previously uploaded signs or sign phrases** by applying i.e. simple copy-paste procedures on pieces of SL utterances or by changing basic components of a sign, using a visual sign editor or the set of HamNoSys notations.
IV) The user may create new SL content by either entering his/her own productions to the system by means of a Kinect devise by exploiting the project's sign language recognition technologies and/or using the sign creation tools and linguistic models also used for editing purposes. In the case of real time input from the part of the user, single signs or sign phrases are performed in dictation style.

V) The user may save, upload and present his/her content preserving his/her anonymity, since performance of sign language content happens by means of a signing avatar exploiting sign animation technologies.

Detailed reporting on end-user evaluation of the DICTA-SIGN Sign-Wiki, will be in the subject of project deliverable D8.2: Evaluation report of Sign-Wiki demonstrator.