A combination of two synchronisation methods to formalise Sign Language animation

Michael Filhol

(1) LIMSI-CNRS, B.P. 133, 91403 Orsay cedex, France

Abstract. The context is Sign Language modelling for synthesis with 3d virtual signers as output. Sign languages convey multi-linear information, hence allow for many synchronisation patterns between the articulators of the body. Addressing the problem that current models usually at best only cover one type of those patterns, and in the wake of the recent description model Zebedee, we introduce the Azalee extension, made to enable the description of any type of synchronisation in Sign Language.

Keywords: Sign Language modelling, synchronisation, Azalee

1 Introduction

In computer animation, more specifically for character animation, the general technique is to establish a set of “key frames” on a timeline, each of which describes a full posture of the character. Then, between every two consecutive key frames a certain interpolation takes place to simulate movement over the frames between the two postures. Iteratively, computer graphists add key frames and refine the interpolations to approach the target movement.

To animate virtual signers—i.e. characters producing Sign Language (SL)—the same technique can be applied. Computer graphists building SL animations “by hand” choose what they feel is the right set of key frames and interpolate between them, smoothening or sharpening paths to eventually come close to gestures that look natural. However, any language processing system needs a formalism to represent what must be signed along the timeline, and the input being a natural language, a linguistically-informed model is necessary.

2 Synchronisation of body articulators in Sign Language

Most SL representation models do not account for time and synchronisation issues as fundamental elements of the SL discourse, and rather define elementary movements and/or sign units that can be concatenated [12, 11, 3]. Some models do promote time to a more essential place in all of their descriptions [9, 4]. Liddell and Johnson have shown that signs (lexical signing units) reveal (rather short) periods when all “features” of the hands align to form intentional postures, between which transitions take place to shift from one to the next, as illustrated in fig. 1a. While this approach slightly differs from the notion of key frame and
interpolation in computer animation (this model being linguistically driven), it does involve fully-specified “key postures” separated by transitions.

We note that all of these models have only really been used, and even designed in the first place, for manual lexical units of SL if not only “frozen signs”, otherwise called “fully lexical signs” [8]. But, in addition to a standard vocabulary of signs, sign languages provide with productive ways of creating new units on the fly [2, 7], which contrast with frozen dictionary entries in that they often enrol more articulators and do not fit the keyframe–transition timeline pattern.

By annotating SL videos\(^1\) with as few articulators as eyebrows, eyegaze and hands (and yet the signing body involving many more), it quickly becomes clear that not all articulators always align in full postures. Many utterances break the set of body articulators into subsets (groups) that seem to act in parallel (fig. 1b). Sometimes the full face acts as a group while the hands are moving; other times a few facial parts take on a different linguistic role and de-synchronise from rest of the face and hands; hands do not always work together; etc.

This makes it impossible only to generate a sequence of body key frames to represent a signed utterance legitimately. While there is good evidence [9] that much of the signing activity does synchronise with key frames, corpus observation reveals that many constructs fall out of the scheme and require a different, more complex synchronisation technique. Our proposition is to combine those two sync schemes into a same language for a better coverage of SL linguistic structures.

---

\(^1\) For this work, we used the French part of the DictaSign corpus [10].
temporal precedence constraints, and each TI specifies what signing takes place within, which together fully specifies the sign utterance. Azalee wraps all this information in an “azalisting” as follows:

```
[[ syncrule1, %% ] List of precedence constraints defining
    ... %% } the layout of TI1, TI2... on the timeline
|| TI1: spec1 %% Sign specification block for TI1
||   ... %% (more blocks, one for each named TI)
]]
```

In the spirit of Zebedee’s necessary and sufficient constraints, TIs are time-arranged with a **minimal** yet **sufficient** set of precedence constraints in order not to over-specify the surface production and allow for any needed flexibility. The syntax and available operators presented here need more corpus analysis to stabilise, but we already propose the few below:

- **interval boundary constraints** for precedence between beginning or end of two different intervals, e.g. `|Gaze = <|Manual` where the beginning of TI **Gaze** is immediately before the beginning of **Manual**;
- **interval boundary ranges** to locate a TI boundary within a range, e.g. `EyeGaze| = |HeadNod ~ WeakHand|>` where **EyeGaze** ends between the beginning of **HeadNod** and a time right after **WeakHand** ends;
- **full interval constraints**, e.g. `Gaze |d| Manual` where **Gaze** is entirely performed during (Allen’s operator [1]) **Manual**.

Each TI specifies its part of the signing activity in its own separate block further down the description, following its name, using either again:

- a key-frame synchronisation of articulators, using a standard zebedescription, i.e. **KEY_POSTURE, TRANSITION** and **SEQ** blocks, as described in [5];
- a TI arrangement there again, using a nested `[[ ]]` azalisting.

The parallel TIs and the possibility of nesting and building sequences recalls the P/C formalism [6], where “partitions” achieve simultaneity of multiple segments and “constituting nodes” bear analogy to key frames. Both P/C and AZee allow for linear and non-linear specification, but the latter allows for **under**-specification (floating boundaries), while the former is fully prescriptive and deterministic. Also, AZee makes overlapping segments easy to specify where P/C needs “null nodes” to pad out the partitions between constituting nodes.

### 4 Conclusion

In this paper, we have introduced the Azalee formalism to describe arbitrary synchronisation patterns, and proposed its combination with the existing Zebedee to form the AZee description language. It re-enables splitting the general signing activity in parts where it is described partially and separately. A consequence is that the different parts of signing taking place in named TIs need to be synchronised, which is achieved by a minimal conjunction of time constraints.
We have described over 2,000 signs and begun clause-level construct description (enumerations, space activation...). It appears that well-established vocabulary units usually call for the key-frame alignment technique, while productive constructs require the more flexible TI arrangement system. But we do not claim analogy between those distinctions. An original aspect of this work precisely is that it does not assume any distinction based on layers of the language prior to its description, rather it provides with a flexible toolkit equally available for all signing activity. For now, we only hypothesise that a time structure tends to simplify as its contents becomes more stable and makes its way into the lexicon.

No implementation is done yet, except for the *AZee* parser. The corresponding solving system still needs to be built, in order to animate virtual signers from an *AZee* input. Though by covering all sorts of synchronisation between articulators, *AZee* accounts for more than the concatenation of lexical units and takes SL modelling beyond the limitation of lexical boundaries of linear grammars. By developing and evaluating *AZee*, we hope to enable valid and precise signing input scripts to virtual signers.

Acknowledgement. The research leading to these results has received funding from the European Community, under grant agreement FP7-231135.

References