

STUDY GROUP 16

Geneva, 26 Jan - 6 Feb 1998

Question: 15/16

SOURCE:¹ Gunnar Hellström, SWEDEN

TITLE: Sign language video test sequences "Irene" and "Rosen"

1 Sign language test sequence "Irene"

This paper and accompanying video sequence presents a video clip with sign language intended to be used as test material for video coding. It contains sign language performed with natural speed.

The sequence is named "Irene" after the signing person. It is in Swedish sign language and copied from Swedish Educational TV. It shows the same head-to-stomach view that is usually used in personal use of videophones for signing. It is recorded in PAL with 25 fps. In the Eibsee Q15/16 meeting it was presented as a file, q11c17A.mpg, MPEG-1 coded in CIF resolution. The intention is to provide it on request on CD, in YCbCr 4:2:0 format in CIF and QCIF 25 fps.

The clip contains a fair amount of fingerspelling, the feature of sign language that from experience is most challenging for videophone communication. Finger movements are so rapid that significant hand positions in some cases are recorded in one frame only, and any framerate reduction below 25 means a loss in language contents.

The clip "Irene", contains a number of eye blinks that are typical grammatical components of sign language used as sentence delimiters. They are short, in many cases only occurring on one or two frames.

From the clips it can clearly be seen that the quality requirements stated before are valid.

- 25 -30 fps is needed for full perception of sign language
- 12.5 - 15 fps gives some usability
- CIF is needed for smooth perception
- QCIF gives some usability

¹ **Contact:** Gunnar Hellström
National Post and Telecom Agency, Sweden
Tel.: +46 751 100 501
Fax: +46 8 556 002 06
E-mail: gunnar.hellstrom@omnitor.se

- Blur in fingers, and face worse than corresponding to QCIF disturbs perception.

Another important factor is the video delay, that should be below 0.4 sec to be good.

Analysis of the frame rate requirement.

Fingerspelling

This is an approximate representation of two fingerspelling sequences in "Irene". The numbers are frame numbers from the beginning of the MPEG version. The letters indicate when the letters are quite clearly formed by the hand. A dash indicates that no clear letter is formed in transition between letters. The first is "Pia Wickman", with the last "a" only visible on the mouth.

"Pia Wickman"	49	-	71	n	317	-
Frame Letter	50	-	72	n	318	v
29 p	51	w	73	n	319	v
30 p	52	w	74	n	320	v
31 p	53	-	75	n	321	-
32 p	54	i	76	n	322	i
33 -	55	-	77	n	323	-
34 -	56	c			324	-
35 i	57	c	...		325	k
36 i	58	c			326	k
37 i	59	-	"Edsviken"		327	k
38 i	60	-			328	-
39 -	61	k	Frame Letter		329	e
40 a	62	k	308 e		330	n
41 a	63	k	309 e		331	n
42 a	64	-	310 -		332	n
43 a	65	-	311 -		333	n
44 a	66	-	312 d		334	n
45 a	67	-	313 s		335	n
46 a	68	m	314 s		336	n
47 -	69	-	315 s			
48 -	70	n	316 s			

There is a grammatical blink in frames 394 and 395.

Among these 18 letters, 5 are clear only on one frame and would risk to be lost at the next logic framerate 12.5 frames per second.

The distribution is:

1 frame - 5 letters

2 frames - 2 letters

3 frames- 4 letters

4 frames- 3 letters

7 frames- 2 letters (ending of phrases)

Conclusion: In this example, the letters within words vary between 1 and 4 frames in time, the frames representing 40 ms each. One frame is in slight dominance. The example is too small for making any real statistical conclusion on, but it can be seen that with this fingerspelling speed, a framerate of 12 would require some guesswork to get what is fingerspelled.

General signing

The last phrase in the clip is signed with signs without fingerspelling, comparable to words.

The phrase is presented here, transcribed sign by sign with the number of frames each sign occupies in parenthesis "SHE(7) TELLS(7) HERSELF(11) HOW(4) SHE (2) FELT(11) EXPERIENCED(13) ADOLESCENCE(16)".

Each sign is a motion, so it is definitely required to reproduce it with a framerate enough to bind the result into moving pictures. (15 fps). The sequence is found between frame 406 and 529 in the MPEG version. The conclusion is that the framerate requirements seem to be slightly less than for fingerspelling. No sign in this sequence was shorter than 2 frames and they did not contain more rapid motion than the fingerspelling.

2Rosen

The sequence named "Rosen" is in American sign language at the Gallaudet University in Washington DC and shows a broader view normally useful for two people in picture. It is recorded in NTSC with 30 fps. The presented file q11c17B.mpg is MPEG-1 coded in resolution 352*240. It is not ready for distribution yet.