

Quality of conversation experience in sign language, lip-reading and text

Andrea Saks

Gunnar Hellström

ITU-T Q.26/16 Accessibility to Multimedia



Some people need to communicate in alternative media

ITU-T

- There have long been efforts to create telecommunications services for people who cannot use voice telephony.
- Picture:
Robert Weitbrecht
Andrew Saks
James C Marsters
with the first successful deaf telecommunications system.
- Just as Alexander Graham Bell's telephone, it is in principle still in operation today.
- Development moves on. New opportunities are created. User's needs to communicate remains.



Time for an accessible replacement of the voice telephone

- o The voice telephone system is inaccessible to many users
- o Video and text additions urgently needed, maintaining global interoperability
- o Possible now through implementation of the Total Conversation standards in new networks and systems: fixed, wireless and mobile.
- o **Let us migrate to accessible telecom systems for all.**
- o Result: Increased markets, increased user satisfaction, fulfilled policy goals.

Specific requirements on media from use with people with disabilities

- o Accessible conversation puts specific requirements on media performance
 - Video for sign language perception is different from video for pleasure
 - Video for lip-reading perception is different from video for information
 - Text for conversation is different from text for side track message exchange
 - Audio for hard-of-hearing language perception is different from Alexander Graham Bell ´s audio



ITU-T Q.26/16 Accessibility of multimedia

ITU-T

- Q.26/16 - Dealing with accessibility of multimedia for people with disabilities.
- Creates own Recommendations and cooperate in accessibility aspects in other Recommendations
- Both human factors and technical aspects
- Conversational services is the main focus area:
 - Video telephony
 - Text telephony
 - Total Conversation



ITU-T

Real Time text conversation

- Text with real-time conversational flow is an often needed component in human interaction
- Often intermixed with Instant Messaging that is important but does not give full feeling of contact

User A	User B
Because then you get a live conversation suitable for a real time call.	Why do you need real-time text transmission? Yes, I see, I can read your thoughts at the moment you express them. No waiting. Good!



Quality requirements on real-time text

- Good conversational experience requires good flow of real-time text.
 - Delay from entry to display:
 - Short delay - good feeling of contact and efficient dialogue
 - Maximum one second for good conversation
 - Two seconds is still usable
 - Three seconds and more cause the traditional problems with delay: Collisions of response and repetition. Stress. "Did he not understand?" . "Am I disconnected?"



Service definitions with real-time text

ITU-T

- Definition of the real-time text for conversation is found in ITU-T F.700 Multimedia Service description framework
- Use of real-time text in services is defined in ITU-T F.703 Multi-media Conversational Services, and F.724 IP Video telephony
 - Text telephony= real-time text and voice
 - Total Conversation= real-time text, video and voice



Occasions when real-time text is important

ITU-T

- o with deaf people, when the users do not lip-read or use sign language
- o with deaf-blind people
- o with signing deaf people when video is not available
- o with hard-of-hearing people
- o with speech-impaired people
- o when environment requires silence
- o when spoken language difficulties arise
- o for all when exactness is important



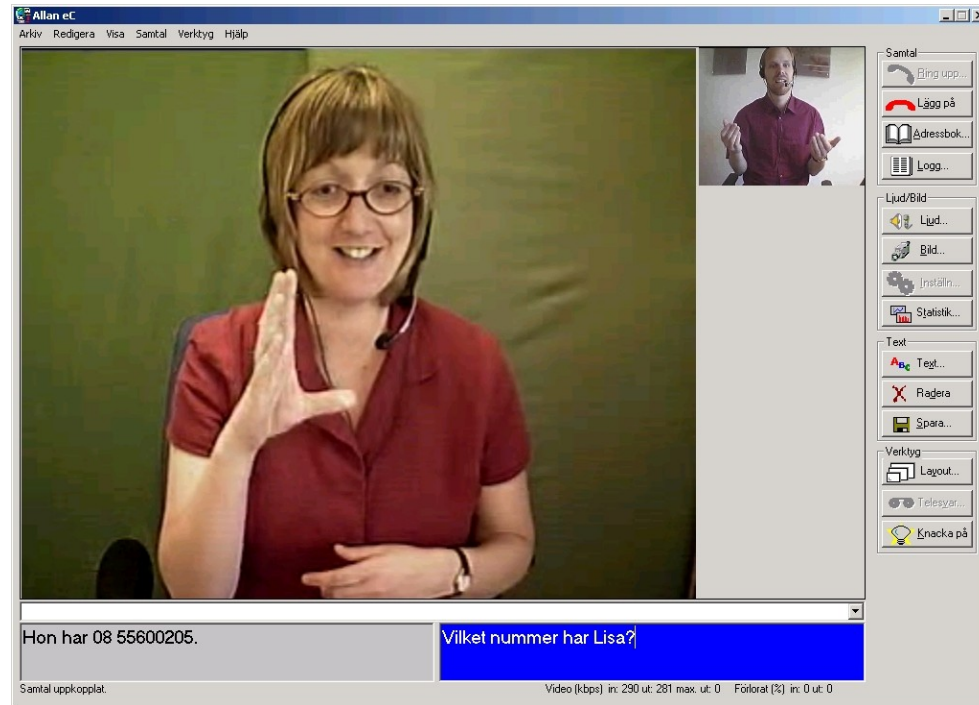
- Smoothness of presentation of real-time text
 - Text may be blocked for more efficient transmission in up to 500 ms intervals.
 - But presentation should be made smoothly character by character.
 - Chunky presentation of text creates a very uneasy feeling by users
 - (research by Omnitor 2005)



Example of a Total Conversation user interface

ITU-T

- o Video, text and voice: standardised
- o Simple extension of the videophone concept
- o Service standardised in ITU-T F.700-series
- o Fulfills a large variety of user needs





Video quality for sign language and lip reading

ITU-T

- o Good sign language and lip reading perception needs 25 frames/s
- o Good sign language and lip reading perception needs 352* 288 pixels
- o Good conversation need delays <400 ms
- o With good video coding this is achievable from 100 kbit/s
- o Too few products perform well today on affordable connections.



Example 1 from fingerspelling

ITU-T

- o Next slide is an example that shows sufficient quality for fingerspelling perception.
- o Fingerspelling of the word "Edsviken" at 25 pictures per second
- o Each letter is represented on at least one picture.
- o Lip reading has similar speed and perception requirements



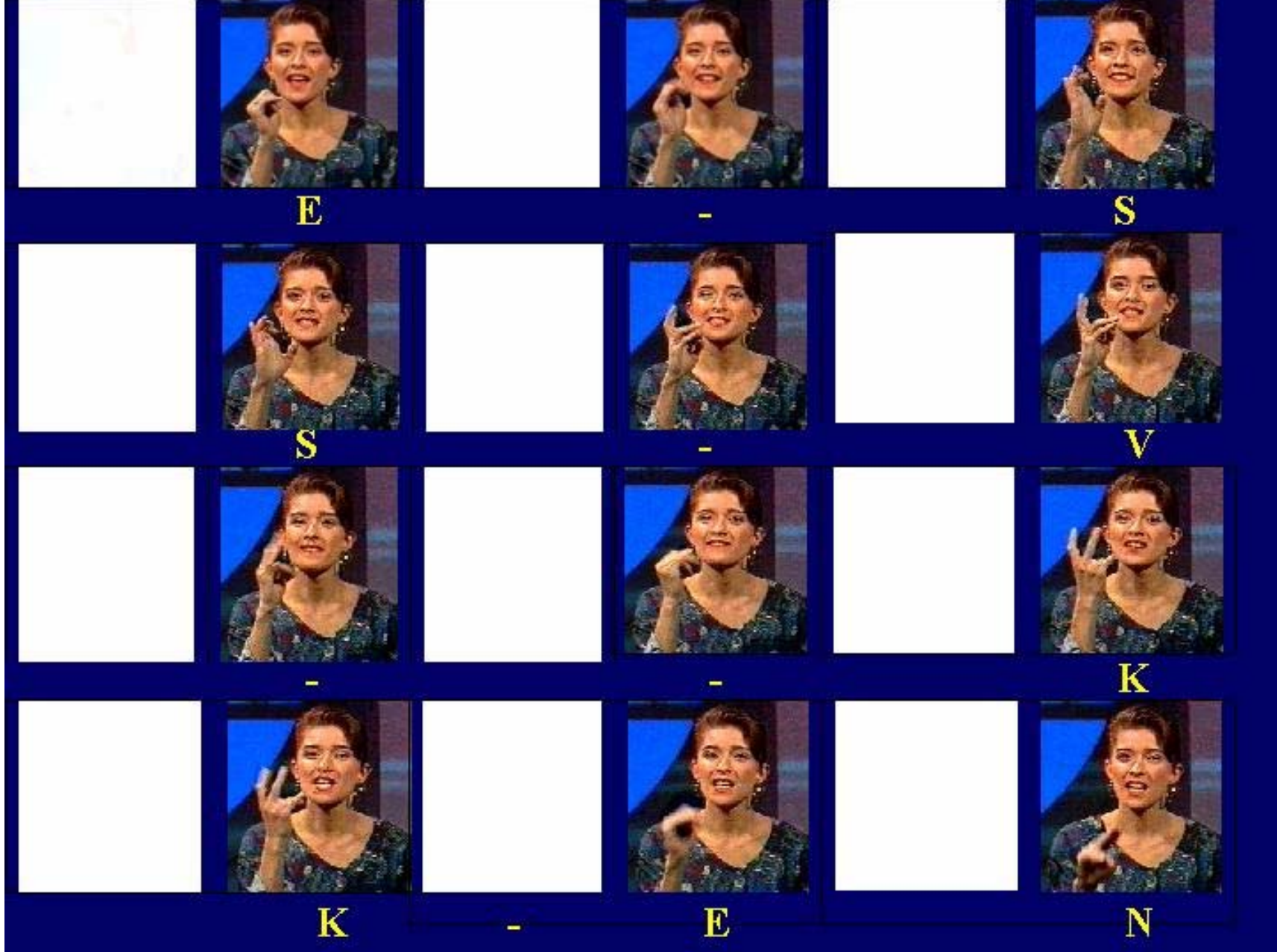
Example: Fingerspelling word "Edsviken" at 25 fps. All characters visible



Example 2 from fingerspelling

ITU-T

- Lowering the frame rate causes loss of language components. Too often done in products.
- Example in next slide: Fingerspelling of the word "Edsviken" at 12 pictures per second
- Result: Only "Esvken" is visible. 2 characters of 8 were lost. Very hard to perceive.
- Similar conclusion for lip reading. Full motion video is important
- Do not be misled by observation that sign language is possible through low frame-rate 3G. It requires severe adaptation by signers.



Example: Fingerspelling word "Edsviken" at 12 fps. 2 of 8 characters lost.



ITU-T

Video quality example

- o Video sample with different resolutions and frame rates.
- o Full motion video at 25 fps is usable, both in CIF and QCIF resolution
- o 12.5 frames per second video is hard to use. Too jerky. Language loss at both CIF and QCIF resolution.

Video quality example



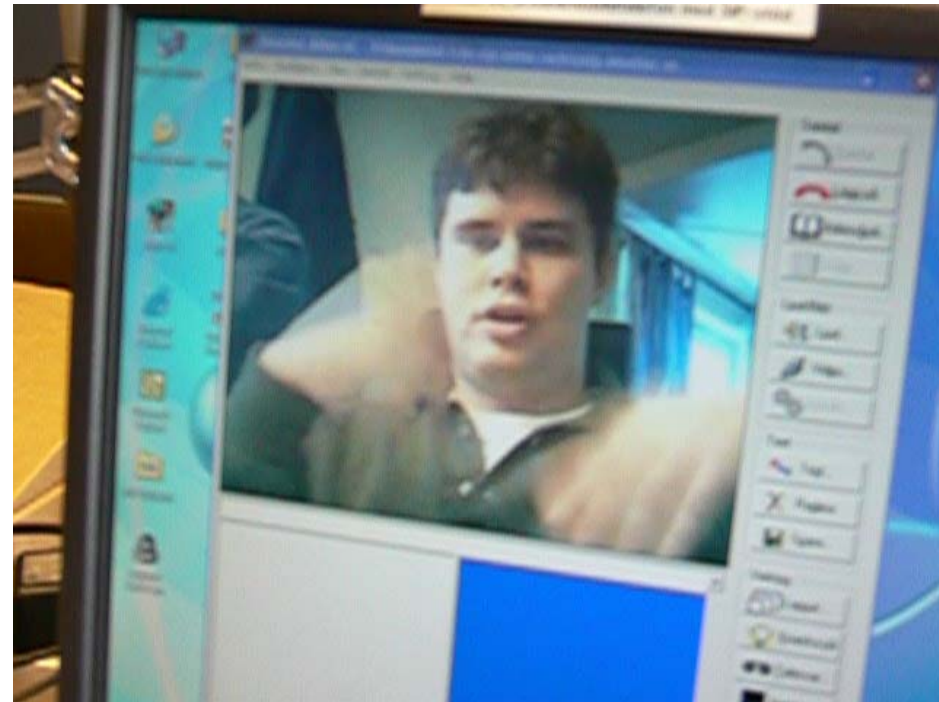


Video requirements documented

ITU-T

- o Video requirements for sign language and lip reading is documented in:
 - ITU-T H-series Supplement 1. "Application profile - Sign language and lip-reading real-time conversation using low bit-rate video communication"
 - Include this reference in all work on video quality.

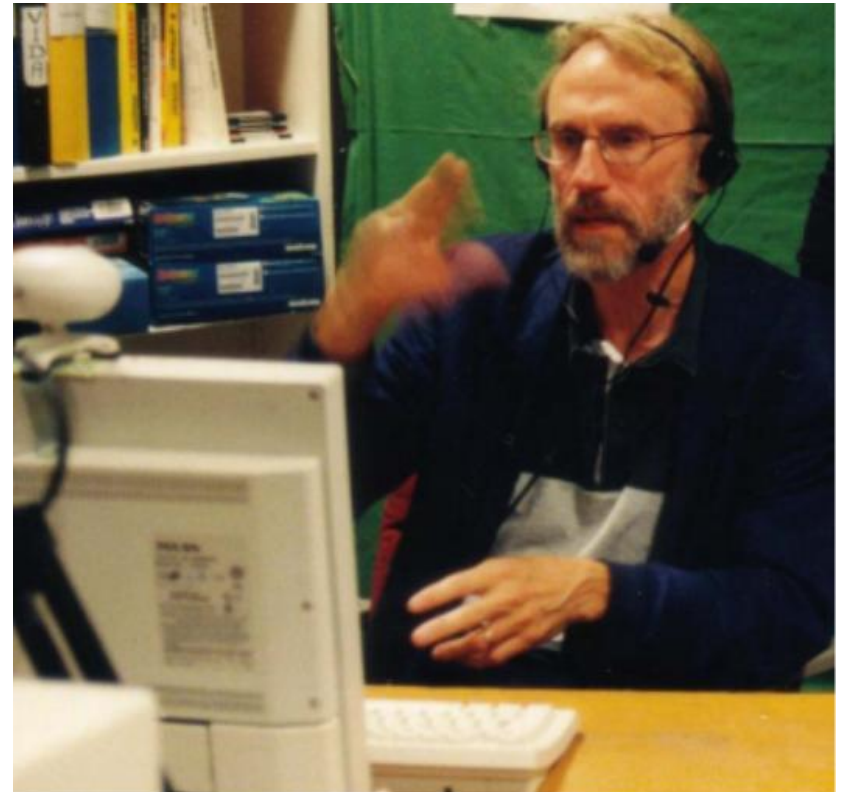
- o Video cameras must use an exposure time of 40 ms or less to produce sharp enough pictures.
- o Longer times create blurry hands and impossible to perceive sign language



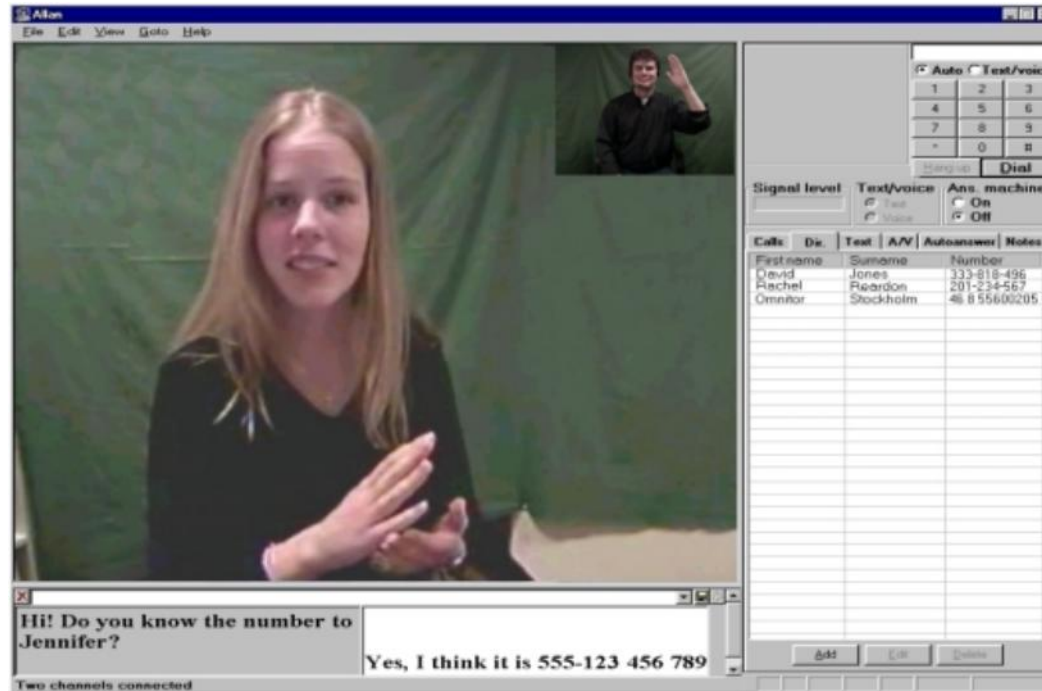
Example from camera with too long exposure time
-a common problem with new web cameras

Example of usability for hard-of-hearing and deaf adults

- Video for lip-reading and acknowledgement
 - Voice for the main conversation
 - Text when stuck
-
- Move to text for the main conversation when the situation calls for it



Example from between deaf sign language users



- Video for sign language for main conversation
- Text important for addresses, numbers, detailed instructions etc.

Good for all – accessibility features add value to mainstream services

- Video for feelings, acknowledgement etc
- Voice for the main dialogue
- Text for addresses, language problems, noise compensation
- ...

(picture from Yoshio Utsumi, General secretary of ITU, and Sylvia Petter, ITU trying Total Conversation demo in UN-days 2001)





ITU-T

Needs in NGN and future work (continued)

- An accessibility checklist is created to ensure accessibility in new standards and designs
- Reminds standardisers on media needs
- Provided to NGN standardisers from Q.26/16



Conclusions

ITU-T

- o The key to accessibility: Provide more media alternatives.
- o Respect the video performance requirements for sign language and lip reading
- o Provide real-time text for fluent interaction
- o Wider audio bands may give hard-of-hearing users a chance
- o Learn more from ITU-T Q.26/16



Slide author: Gunnar Hellström, Omnitor

e-mail gunnar.hellstrom@omnitor.se

Tel: +46 708 204 288

Fax: +46 8 556 002 06

www.omnitor.se

Andrea Saks

e-mail. asaks@waitrose.com