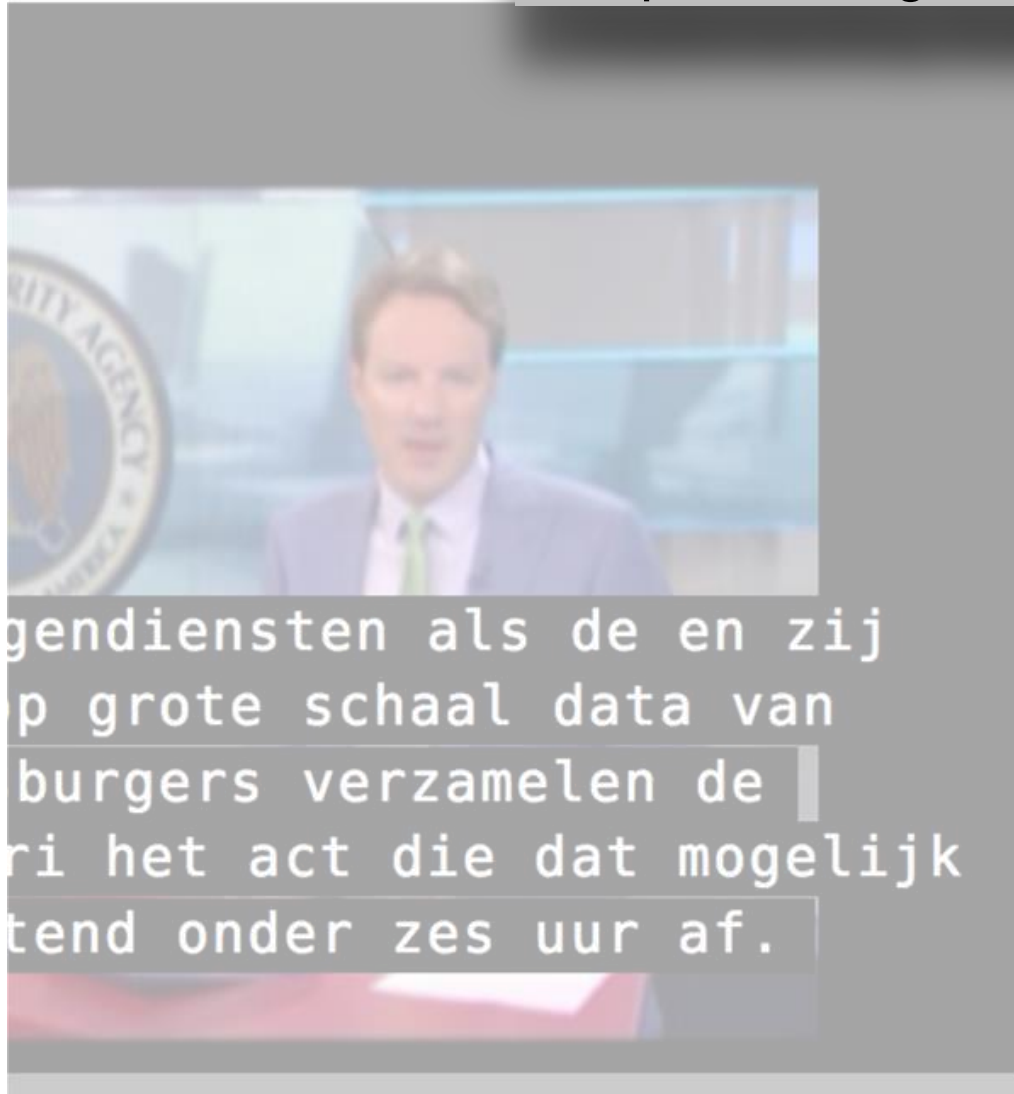


Automatic Subtitling

Implementing Automation in Subtitle Production



journalisten me voor een verg
gemaakt van een Uitgebreide vers
ook zeggen ze dat de satellie
Russische defensie zijn

Start 00:03:58:02 Length 00:00:13:08 End

Amerikaanse inlichtingendienst
kunnen niet langer op grote
Amerikaanse staatsburgers
anti-terreur wet de Petri het ad
maakt, liep vanochtend onde

Erik Buitinga
NPO Subtitling, The Netherlands

Automatic Subtitling

Overview

NPO implemented a *Semi-automatic* subtitle production process using speech and language technology that can be plugged into an automated workflow for subtitling

Result of a five year NPO R&D project
in collaboration with software developer Optispeech

An Optispeech product called eCaption, 'computer assisted subtitling', was developed

February 2014 implementation of eCaption system at NPO
October 2014 implementation of eCaption 'webeditor' at NPO

By now 400+ programs have been made accessible
using this technology

Automatic Subtitling

Project challenges

- It hasn't been done before! Previous attempts to set speech technology to work in subtitling were unsatisfactory
- Automation of a professional specialism is risky business
- Subtitle text contains more than just dialogue
- Broadcast content audio contains a lot more than just dialogue
- Broadcast dialogue is not 'clean' speech
- Outcome of a semi-automatic process has to meet the standards applied to traditional subtitling in some way
- Businesscase dictated that the solution should yield a 20% reduction of subtitle production time for 50% of pre-prepared broadcast content
- We know about subtitling, not about speech&language technology or development!

Automatic Subtitling

Lessons learned from previous endeavors

- Fully automatic subtitling is not possible (yet)
- Automated system should not operate alongside the subtitling process, but should integrate it
- Whatever the process, it must integrate text editing to produce proper subtitles in the end
- You'll need a suitable platform for this
- The process must handle non-dialogue text intelligently
- System must be self-reporting
- System must allow an editor to work through the videofile only once
- System must allow for standard subtitle operations on problematic segments

Automatic Subtitling

The eCaption project

- Started in 2010 with the development of a Dutch language model based on broadcast speech by LIMSI/CNRS (France)
- Some 500 hours of video with associated subtitle files were processed for a basic language model
- 100.000-plus subtitle files were processed for extended vocabulary
- 10 hours of video where transcribed (verbatim) for system calibration purposes
- A basic interpunction model was developed
- Speech to text/speech alignment system as a result
- Separate software module was developed by SempreSole (Netherlands) for XML-fileprocessing and formatting timed text into subtitles

Automatic Subtitling

Process philosophy

- Speech recognition technologies and system automation can facilitate subtitle editors in their jobs only if subtitle editors can help the applied technologies perform better
- So, automatically retrieved data and user generated data are integrated to reinforce each other's validity
- Process is dedicated to subtitling only
- *eCaption subtitles are produced in a 2-step process, every step consists of automatic data extraction or processing followed by data verification/enrichment by an editor*

Automatic Subtitling


Save | Cancel | Wiki | [Give feedback](#)

ECAPTION
SUBTITLE WEB EDITOR COMPUTER ASSISTED SUBTITLING


Erik Buitinga

NOS_JOURNAAL_- WON01464253_TRANSCRIPT_ECAPTION

Start **00:03:47:03** Length **00:00:09:00** End **00:03:56:03** ✓



Amerikaanse inlichtingendiensten als de en zij kunnen niet langer op grote schaal data van Amerikaanse staatsburgers verzamelen de anti-terreur wet de Petri het act die dat mogelijk maakt, liep vanochtend onder zes uur af.

00:04:05:02  00:12:14:00

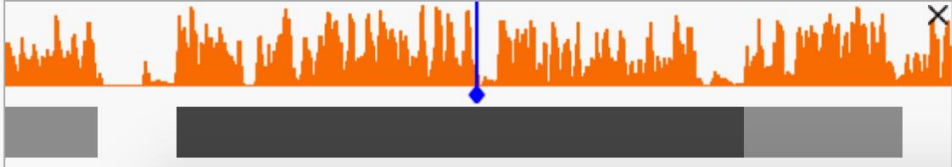
journalisten me voor een vergelijking gebruik gemaakt van een Uitgebreide versie van Google Earth ook zeggen ze dat de satellietbeelden van de Russische defensie zijn bewerkt.

Start **00:03:58:02** Length **00:00:13:08** End **00:04:11:11** CPS **16** ✓

Amerikaanse inlichtingendiensten als de en zij kunnen niet langer op grote schaal data van Amerikaanse staatsburgers verzamelen de anti-terreur wet de Petri het act die dat mogelijk maakt, liep vanochtend onder zes uur af.

Start **00:04:11:11** Length **00:00:03:16** End **00:04:15:02** ✓

Amerikaanse Senaat kon het niet eens worden over een verlenging van de wet.



Automatic Subtitling

The eCaption process

- **ingest** - automatic ingest of video, audio, text (post production script), translated subtitle file (indicating 'no-go areas' for the s2t-process), metadata (broadcast filename, on air date etc.)
- **analysis** - speech to text process produces a timed transcription (filtered or non-filtered) of dialogue and a system report is produced
- **editing** - analysis result is presented in web editor as timed text; editor adds/corrects missing/misinterpreted words, adds missing timestamps, non-dialogue text and specially formatted text
- **alignment** - enriched data from editing process is aligned to video timecode, formatted into subtitles according to user settings and a system report is produced
- **verification** - editor checks verification markers as reported by the system and makes necessary adjustments
- **finalization and delivery** - final result is converted to a subtitle file of choice and exported from the eCaption environment

Automatic Subtitling

Benefits

Depending on audio- and speaker quality of broadcast content

- *Automatic* transcription of dialogue (full or filtered transcription)
- *Automatic* word alignment
- Insertion of timed non-dialogue text and application of special formatting rules while editing
- *Automatic* formatting into aligned subtitles
- Error reports on alignment and formatting process
- Reports on system accuracy, percentage of speech and estimated production time
- Easy adjustment of settings
- Integrates with an *automated* production workflow
- No need for dedicated MAM-system
- Also supports traditional subtitling
- It's easy to learn/crowdsourced subtitling

Automatic Subtitling

Does it beat traditional subtitling?

- So far there's a reduction of production time for programs with relatively clean audio and relatively good speech when the process is run by a traditionally trained editor. This type of program definitely doesn't make up 50% of our pre-prepared broadcast material.
- New staff, especially trained to work with eCaption, performs much better. On average they'll reach production values of 10 to 13 minutes of texted video per hour (compared to 7 minutes of subtitled video by a traditional editor) regardless of the quality of speech and the speech/noise ratio of the broadcast content.
- Time needed for verification should be added to this: verification time will increase when speech quality is low and when speech/noise ratio is tilted towards noise
- In some cases subtitles have been obtained almost automatically!
- eCaption subtitle files differ from traditionally produced ones: Typically, they'll contain significantly more words and therefore generate higher reading speeds. Also, its subtitle lay out differs.