

Easing Transcripts for MOOC Videos with an ASR (Automated Speech Recognition) System

Carlos Turró, Jorge Civera and Jaime Busquets
Universitat Politècnica de València



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA









The result of not having a screwdriver

- Pain
- Frustration
- Select a different tool



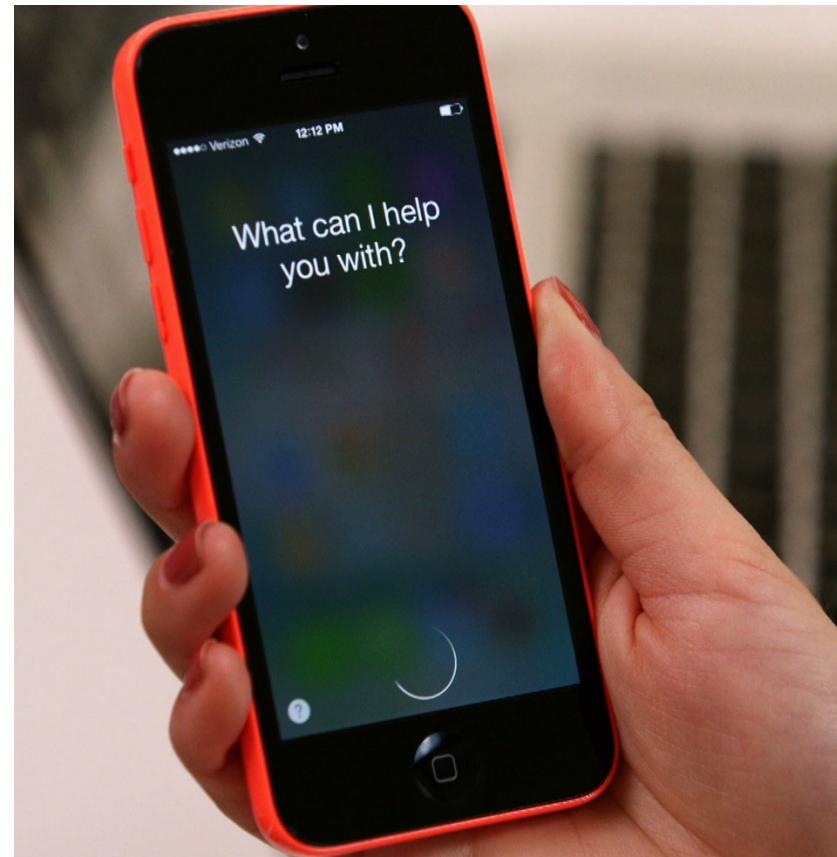
How can I transcribe a video?

- Manually transcribing a video takes 10 times the length of the video (RTF)
- Boring
- It's worse if you don't know about the topic of the video



Automated Speech Recognition (ASR)

- How good is it?
- Will it recognize my special words?
- Will it really help me?



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

UPValenciaX MOOCs - Transcribing

<https://media.upv.es/player/?id=be0c7738-039d-9445-8237-8b85f37cd303&autoplay=true>

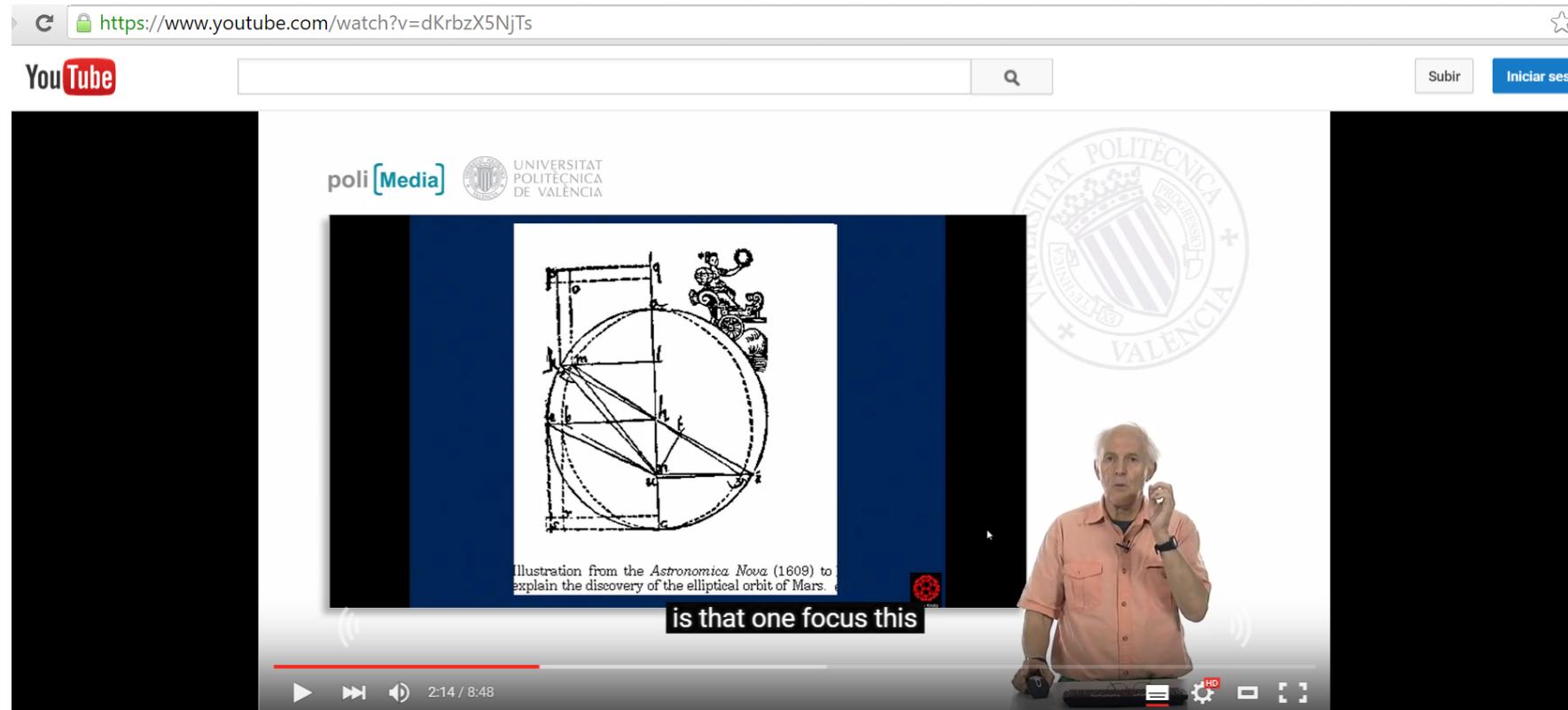
The screenshot shows a video player interface. At the top left is a play button icon. Next to it are the logos for poli[Media] and the Universitat Politècnica de València. The main content area features a historical astronomical illustration from 'Astronomica Nova' (1609) explaining the discovery of the elliptical orbit of Mars. The illustration depicts a celestial model with various circles and a figure. Below the illustration is a caption: 'Illustration from the *Astronomica Nova* (1609) to explain the discovery of the elliptical orbit of Mars.' A small red circular logo is at the bottom right of the illustration. To the right of the illustration, a man with white hair, wearing an orange shirt, is speaking. He is positioned in front of a large watermark of the Universitat Politècnica de València seal. A dark blue bar at the bottom contains the text: 'about the motions of the planets and the particular one that was most important was this one'.



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

<https://media.upv.es/?id=b444d12e-db23-9a4f-9b3b-d1d9275d4cb4>

UPValenciaX MOOCs - Transcribing



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

<https://www.youtube.com/watch?v=dKrbzX5NjTs>

UPValenciaX MOOCs - Transcribing

The screenshot shows a web browser displaying the UPValenciaX MOOCs page on the edX platform. The URL in the address bar is <https://www.edx.org/school/upvalenciam>. The page features the edX logo and navigation links for Courses, How It Works, Schools & Partners, and About. A search bar contains the placeholder "I want to learn about...". On the right, there are "Sign In" and "Register" buttons. Below the navigation, there is a large image of a park-like setting with people sitting on the grass under trees. To the right of the image is the logo of the Universitat Politècnica de València (UPV) and its name in Spanish. At the bottom left, the text "UPValenciaX" is displayed, followed by "Free online courses from Universitat Politècnica de Valencia". At the bottom right, there is a link "Back to schools and partners".

30 MOOC courses



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

UPValenciaX MOOCs -Transcribing



UPValenciaX MOOCs – Transcribing



Transcription and Translation Platform

- Post-editing web interface (in HTML5)

The screenshot shows a video player interface. At the top, it says "State University of New York". Below that is a list of statistics:

- 64 campuses
 - 30 Community Colleges
 - 8 Technology Schools
 - 13 University Colleges
 - 13 Doctoral Institutions
- 468,000 students
- 88,000 faculty and staff

Next to the list is a map of the State University of New York system, showing various campuses across New York state. Below the map is the SUNY logo. To the right of the map is a circular seal for the Universitat Politècnica de València. At the bottom of the video player is a control bar with icons for navigation, volume, and other media controls, along with a timestamp "00:20 / 07:43".

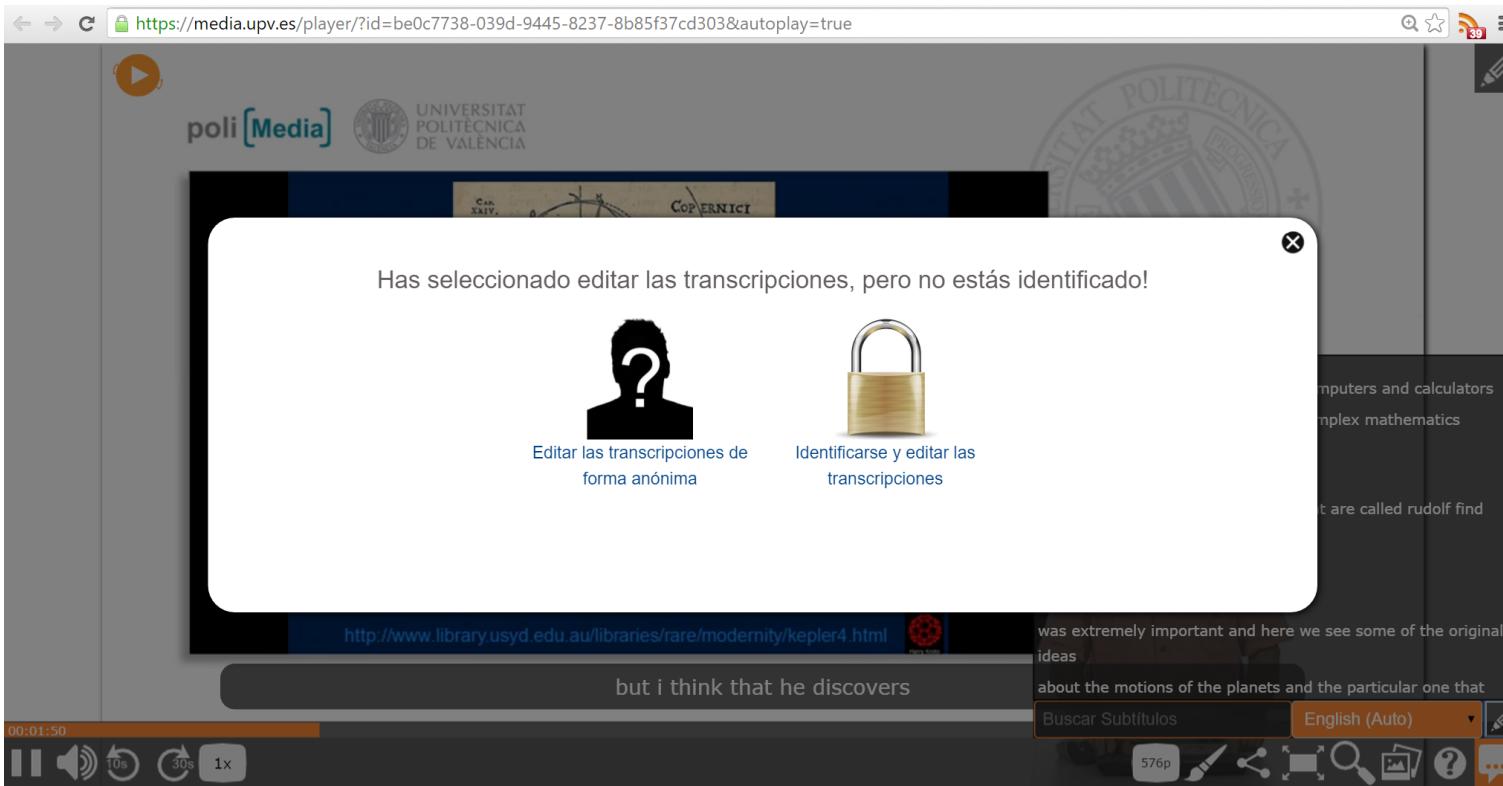
The screenshot shows a transcription interface with a large text area containing a speech transcript. The transcript reads:

my name is kim scales so and i'm from the state university of new york and
today i'm going to talk to you about the way that we have used
a self-assessment using the date tools for organizational development
a little bit about the university of york first we are sixty four campus system with over four hundred and
sixty thousand students and eighty eight thousand faculty instead
the organization that i am is the city center for professional development and mirrors central organization providing training
and professional development for faculty and staff on campuses
what i'm going to talk about in terms of the case study is based
on how to use the self-assessment for organizational development
with a variety of organizations i've conducted self-assessment for my own organization three different times and that's across two different
organizations
my also facilitated this process once with essential distance-learning organisation
and i have been involved in facilitating many workshops and conference sessions over the last three years



Crowdsourcing

- We are crowdsourcing the on-campus courses using our own Paella video player.



UNIVERSITAT
POLITECNICA
DE VALÈNCIA

How to get good transcription quality

- Transcription systems learn to transcribe from examples
 - At least 50 hours of videos (audio) in the source language previously transcribed to learn the acoustic model
 - Texts in millions of words to learn the language model

Language	Videos (hours)	Text (Mwords)
Dutch	532	628
English	620	464000
Estonian	130	410
French	88	1800
German	36	135
Portuguese	54	573
Italian	54	868
Slovene	27	224
Spanish	128	654



How to get good transcription quality (II)

- Adaptation of transcription systems to the specific videos is key for high accuracy
 - Availability of videos manually transcribed with similar acoustic conditions
 - Availability of text resources related to the video in question
 - Title is used to retrieve related documents
 - Slides contain most of the special words used by the lecturer
 - Documents: text content from the course, additional text resources (bibliography)
- Sound quality of the video has a direct relationship with quality
 - No noise, no background music, please



Try yourself

The screenshot shows the homepage of the MLLP website. At the top, there is a navigation bar with links for Home, Members, Research, Try our tools, and Contact. Below the navigation bar is a large banner featuring a person's hands typing on a laptop keyboard. Overlaid on the banner is the text "Try our transcription and translation technologies" and a description: "The MLLP offers state-of-the-art automatic transcription and translation services for large video repositories." Two buttons are visible on the banner: "Free trial" and "How does it work?". The URL "www.mllp.upv.es" is visible in the browser's address bar at the top left.

<http://mllp.upv.es>



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

Our next step



Translations !!



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

Conclusions

- ASR technology is enough mature to help a lot in captioning
- However, there should be a review phase
- Quality can be enhanced by providing transcribed videos
- At UP Valencia we got transcribed our 30 MOOC courses with 3x TA cost ☺



Thanks!

Questions?



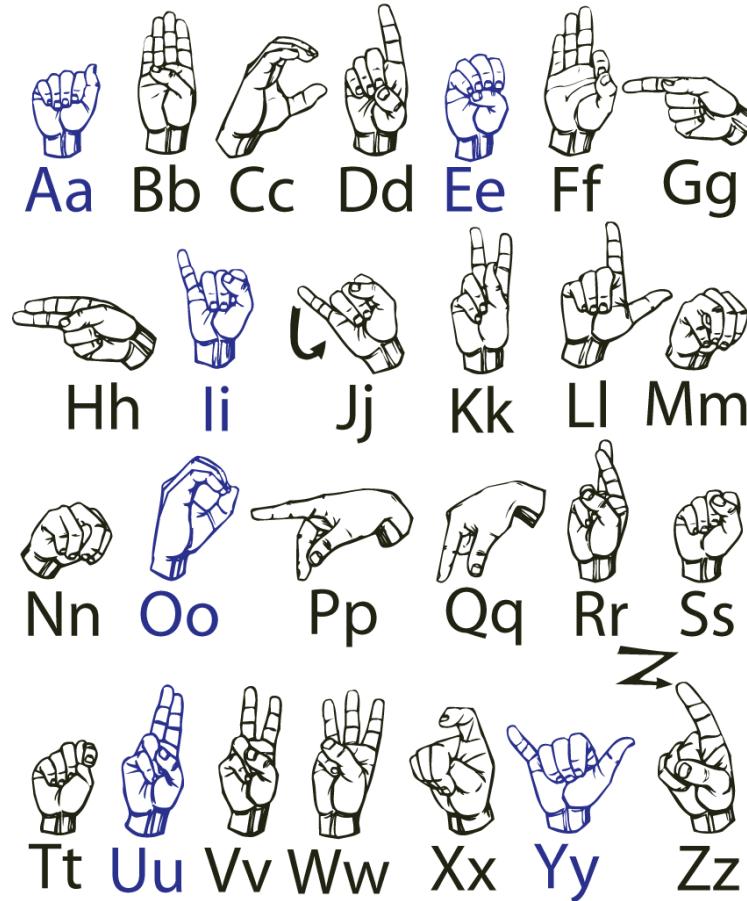
UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

Why transcription of MOOC video files?

- Accessibility



Why transcription of MOOC video files?

- Accessibility
- Searching into a video file
- Searching into a video repository
- Topic identification
- ...and much more



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

Measuring Quality: Word Error rate

$$WER = \frac{S + D + I}{N}$$

Where

S is the number of word substitutions,

D is the number of word deletions,

I is the number of word insertions,

N is the number of words in the reference text



Measuring Quality: Word Error Rate

Language	WER
English	20.8
Dutch	24.5
Italian	17.7
Spanish	14.4
Estonian	27.1
French	22.7



Attributions

- Fingerspelling & tools Wikipedia
- Bored <https://www.flickr.com/photos/left-hand/3132070992/>
- Siri <https://www.flickr.com/photos/smemon/8070397213/>

