

Importing YouTube Data and Media in MAXQDA 2018 - Addressing Data Access Hurdles

Matthew Loxton

Abstract: Using MAXQDA 2018.2, you can directly import YouTube comments and manually created transcriptions, but cannot import the video file or transcripts created automatically by YouTube. This post walks you through a way to acquire and download video files and import transcriptions created automatically by YouTube.

Keywords: Acquiring transcripts, computer assisted qualitative analysis, import, automated transcription, timestamps, YouTube

Table of Contents

1. Introduction
2. Hurdles with YouTube Data Access
3. The Media Hurdle
 - 3.1. Acquiring the Media File (Ask the Owner)
 - 3.2. Acquiring the Media File (YouTube Premium)
 - 3.3. Acquiring the Media File (App)
4. The Transcription Hurdle
 - 4.1. Acquiring the Transcript
5. Conclusions
6. About the Author

1. Introduction

Using MAXQDA 2018.2, you can directly import YouTube comments and manually created transcriptions, but cannot import the video file or transcripts created automatically by YouTube. This post walks you through a way I used to acquire and download the YouTube video file (MP4), and import transcriptions created automatically by YouTube.

2. Hurdles with YouTube Data Access

For a researcher, there are many non-commercial YouTube videos that would be very good sources of data. However, it is often not possible to just download the video as a media file to use within MAXQDA. Additionally, in many YouTube videos, the transcriptions are automatically generated by Google Machine Learning, and cannot be automatically downloaded by MAXQDA.

I attempted to find ways to bridge these hurdles, and this blog describes what limited but useful success I had.

3. The Media Hurdle

There are three main ways to acquire the video media:

1. Ask the author or channel owner for the media file;
2. buy the YouTube premium membership; or
3. access the video directly on the YouTube server location and save it to your PC.

3.1. Acquiring the Media File (Ask the Owner)

I had some success simply asking the authors or the YouTube Channel owner for a copy of their video. In approximately half of the times I asked, the end result was receipt of an MP4 media file. The chief obstacles after establishing contact were that the media files were too large to be simply emailed, and required additional steps of the author placing them on a sharing site such as Drop-Box or Google Drive. Each added complication resulted in a steep drop-off of success.

3.2. Acquiring the Media File (YouTube Premium)

YouTube Premium allows the user to save an encrypted version of the media file to their mobile device that can be played back without access to the Internet. This makes the researcher less vulnerable to risks of the video being taken off the internet, or an Internet connectivity disruption. Although the media file is saved to the device, it will nonetheless expire, can only be viewed using the YouTube app, and cannot be played on a PC or within MAXQDA.

3.3. Acquiring the Media File (App)

There are many commercial off the shelf tools and applications that can be used to download YouTube videos, and I will cover one free open-source application. The VideoLAN non-profit describes its VLC product as a “free and open source cross-platform multimedia player and framework that plays most multimedia files, and various streaming protocols.”

After downloading and installing VLC, open the YouTube browser for the target media (See Figure 1) and copy the URL (1).

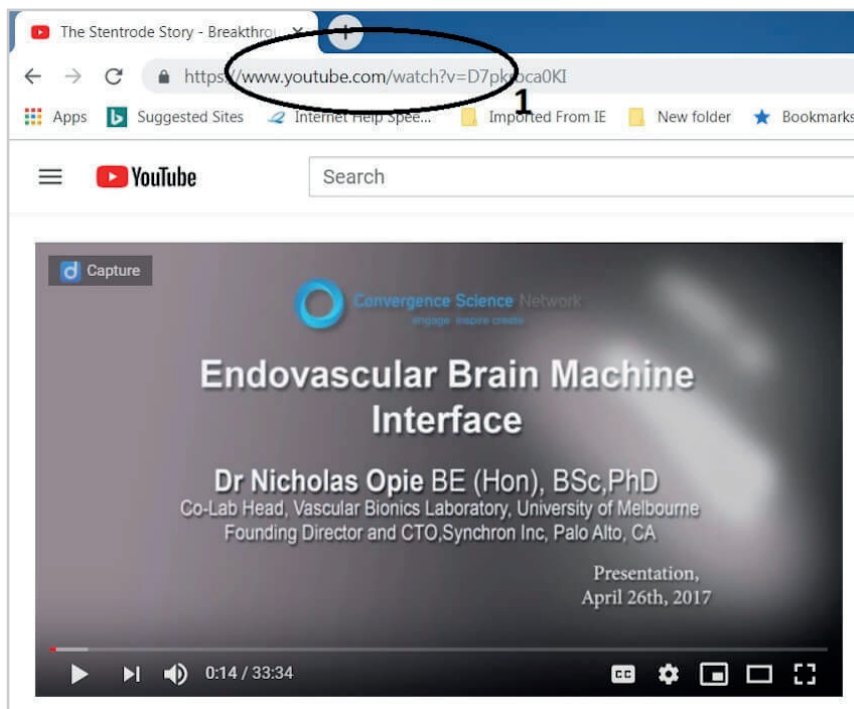


Figure 1 YouTube Browser and Target URL

Open VLC and click on “Media”, select “Open Network Stream” and paste the full YouTube URL into the network textbox (Figure 2 Item 2), and click “Play” (3).

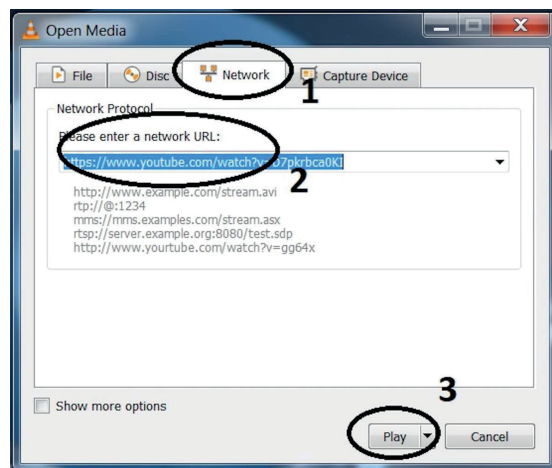


Figure 2 VLC Open Media Dialogue

Once you have verified that the correct YouTube video is playing, click on "Tools", and select "Codec Information" from the VLC dropdown. In the resulting Media Information screen (Figure 3) verify that you are in the Codec panel (1), and then copy the long URL from the "Location" text-box. This is the full path to the YouTube media.

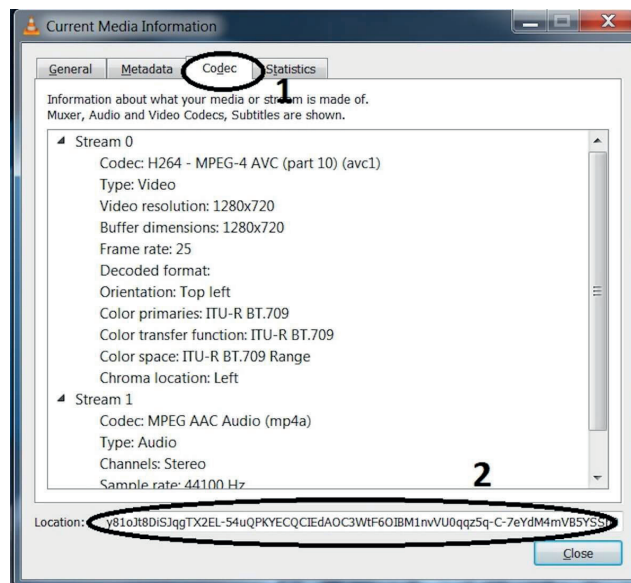


Figure 3 VLC Media Information – Codec

Paste this long URL into your browser and once the video begins to play, right-click on the video and select "Save video as". Save the media file to a memorable location. Once it has downloaded, you may wish to double-click it to verify that the MP4 file plays and is the video you wanted.

4. The Transcription Hurdle

If the YouTube video has a manually-created transcription, you will see something like the image shown in Figure 4. The import of YouTube data is fairly simple and uses the MAXQDA YouTube Data function available in the Import section.

You can do this via the following steps:

1. Navigate to the YouTube video with a browser
2. Check that a transcription exists
3. Select the URL and copy it
4. In MAXQDA, click on **Import**
5. Click **YouTube Data**
6. Observe that correct url has populated the URL field in the import popup or paste it in
7. Observe selection for importing comments

In the **MAXQDA manual**, the image of the dialog box shows a section for **Import transcript/closed caption** (1) and an option to **Link existing video file to transcript** (2)

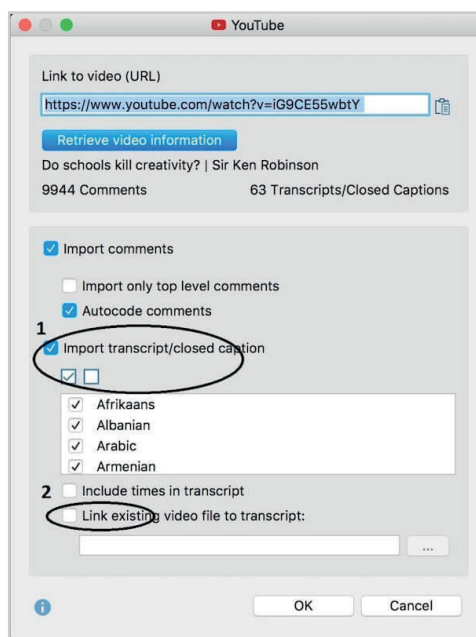


Figure 4 Import Transcript

However, if the transcription was generated automatically by YouTube, you will see a slightly shorter screen, with no option to download the transcription or link to media.

For example, the YouTube video I wanted had an automatically generated transcription file, and following the same steps instead yields Figure 5.

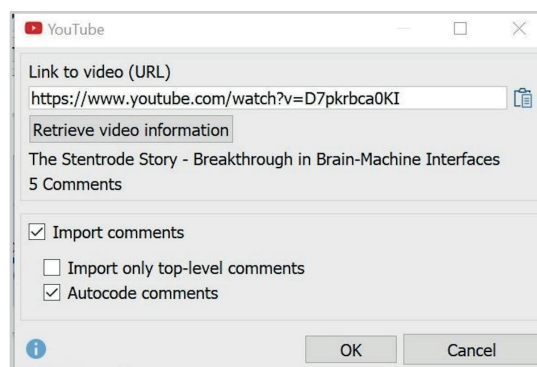


Figure 5 YouTube Import with Automated Transcription

To get past this obstacle and have a linked transcription and media file in MAXQDA, you can manually acquire the media file (see above) and transcription with timestamps, and import and link them in MAXQDA.

4.1. Acquiring the Transcript

To download the automated YouTube transcript, you can manually copy it and edit the timestamps in a way that it can be imported into MAXQDA that allows it to be linked to the media.

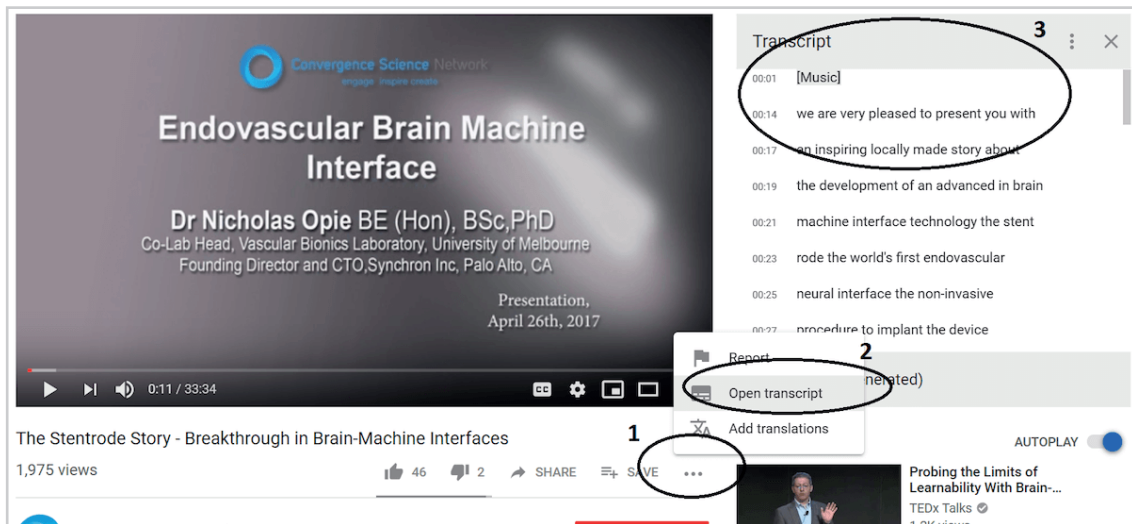


Figure 6 Acquiring the Transcript

1. In the YouTube browser screen, open up the automated transcript by clicking on the "... " (Figure 6 Item 1) icon under the viewing pane, and select **Open Transcript** (2).

2. In the transcript panel (3) select the first few characters, scroll to the bottom, shift-click at the end to select the entire transcript and copy.

Notes:

- If the timestamps are not visible, click the "... " icon and select Toggle Timestamps
- I have found this manual selection method works best because using a "select all" shortcut like **Ctrl-A** will harvest things like page headers and footers that you do not want.

3. Open Microsoft Excel or other spreadsheet

4. Select the first column, and format as text (if you don't do this, it may interpret the timestamps and complicate the next steps).

5. Paste the transcription into the first column

6. You may see something like the following, with text and the timestamps in subsequent rows.

```
[Music]
00:14
we are very pleased to present you with
00:17
an inspiring locally made story about
```

00:19

the development of an advanced brain

00:21

Note: If you import this, MAXQDA will not yet be able to recognize the timestamps, so these need to be formatted first.

7. To format the timestamps, I used the following Excel equation in column 2: `=IF(ISNUMBER(VALUE(LEFT(A1,1))),"#00:"&A1&"#",A1)`
What this does is look at the cell in col 1, evaluate if the value of the first character is numeric, and if it is, it adds a hash and a 0 to the front of the timestamp, and adds a hash to the end. If the first character isn't numeric, we assume it is the transcription text, and just copy it into col 2.
8. Select col 2
9. Open a word document (or any word processing app), and paste the column as plain text.
10. Save the document in a memorable location.
11. In MAXQDA, click on **Import > Transcripts with Timestamps**, and select the saved word document.
12. When the dialogue box prompts you, select the YouTube media file, and browse to the file.
13. MAXQDA will import the transcript and you will see a popup dialog box inviting you to link the media file.
14. Navigate to the media file you saved above, and press **OK**

5. Conclusions

Although there are legal and practical barriers to automatically downloading YouTube media and machine-generated transcripts, there are somewhat straightforward manual methods that can enable the researcher to achieve the same thing. These methods may not age well if YouTube or VLC change their product features, but in the interim, they can be a practical means of having transcripts that are linked to video media for analysis.

About the Author

Matthew Loxton is a Principal Analyst at Whitney, Bradley, and Brown Inc. focused on healthcare improvement, serves on the board of directors of the Blue Faery Liver Cancer Association, and holds a master's degree in KM from the University of Canberra. Matthew is the founder of the Monitoring & Evaluation, Quality Assurance, and Process Improvement (MEQAPI) organization, and regularly blogs for Physician's Weekly. Matthew is active on social media related to healthcare improvement and hosts the weekly #MEQAPI chat.