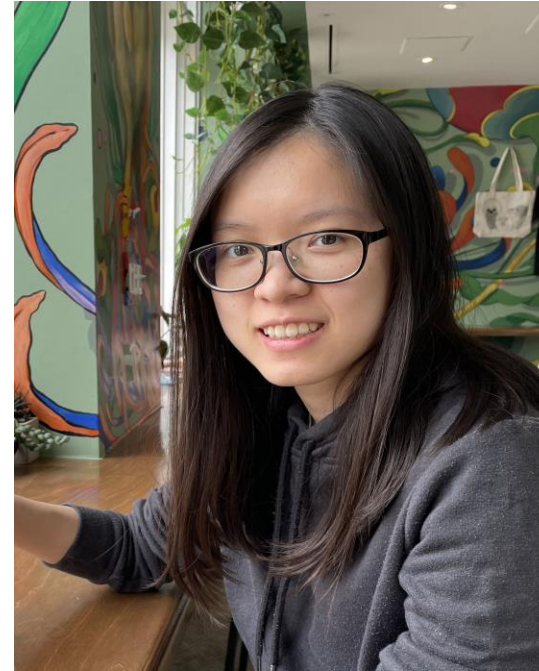


Critical Decisions to Make in Subtitling and Dubbing Projects – For Project Managers

Abby Chung & George Zhao – VideoLocalize – GALA Webinar 2021



George Zhao
President & CEO
george@videolocalize.com



Abby Chung
Localization Engineer
teamto@videolocalize.com

VideoLocalize – What we do

- Specializes in video localization for corporate videos
- VideoLocalize is a venture started by Boffin in 2017
- A Video Localization Management System and Dubbing Automation tool
- Winner of multiple Innovation Awards (TAUS, LocWorld)



TAUS Game Changer Innovation 2016



LocWorld Barcelona –
Process Innovation Challenge 2017



TAUS Game Changer Innovation 2017

What we'll cover today

- Subtitles
- On-Screen Text (OST)
- UI Text
- Dubbing – Timed or Untimed Recording
- Dubbing – Human or AI Voice Over
- Q & A

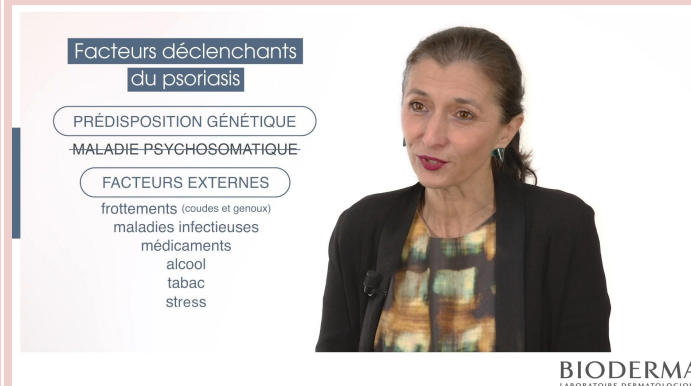
Subtitle / On-Screen Text / UI Text

Subtitles



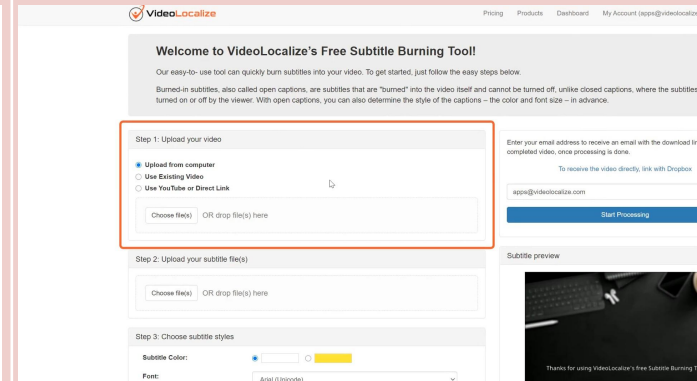
Captions displayed at the bottom of the screen that translate or transcribe the dialogue or narrative.

On-Screen Text (OST)



Text that shows important information or topics of the video.

UI Text



User Interface text, usually includes control labels and static text.

Subtitling Questions

- Which type of script to translate: script with or without timecodes?
- How does the translation process affect translation memory and quality?
- What if the video has on-screen text or UI text?

Types of Transcription

Transcription without timecodes

And sometimes what's important about hearing a message is hearing it more than once because we hear it in unique ways. We're in a new place today, we're in a new season. And so I'm going to give you some strategies and I'm going to share with you kind of where our focus is for our time together.

You're going to learn how to develop a courageous or optimal mindset. So, I want you to touch your head quick. So, an optimal mindset, a courageous mindset, one that is working to your collective advantage. It's like you are at your best.

Then the second area of focus that we're going to cover is we're going to discover a few EQ strengthening techniques and that will support you to navigate each opportunity with mindfulness, confidence, compassion, and enthusiasm. Okay, so mindfulness, confidence, compassion, and enthusiasm, jazz hands.

Okay? All right, so, this is our focus. This is the direction in which we are going to travel today and let's take a quick peak. So, we talked about this at conference, emotional intelligence, EI, emotional intelligence. This is how our ability to read people, to have some self-awareness, to be able to navigate conversations so they're smooth and people leave that conversation feeling like you get me, you understand me, or you shared with me just what I wanted to hear today.

This really resonated with me. When we have this high emotional intelligence, we're able to be agile in our communication. And this is the definition we shared with you, being intelligent about our emotions. So, being intelligent about our emotions and then using emotions to unlock our intelligence. So, being intelligent about our emotions and using our emotions to unlock intelligence.

Transcription with timecodes

```
1
00:00:03,937 --> 00:00:06,106
Security is a full stack exercise:

2
00:00:06,106 --> 00:00:07,957
from the hardware at the base

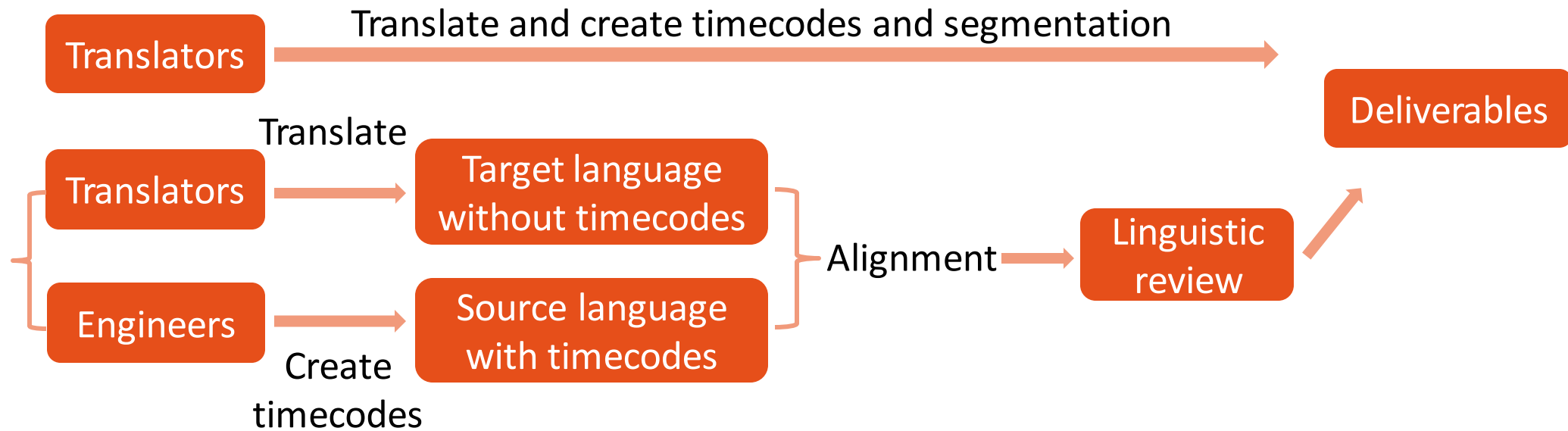
3
00:00:07,957 --> 00:00:10,777
to the operating
systems that might run upon it,

4
00:00:10,777 --> 00:00:13,747
to the transport protocols
that carry bits back and forth

5
00:00:13,747 --> 00:00:16,566
among different nodes up
through the software running

6
00:00:16,566 --> 00:00:19,936
on those operating systems
and of course, people,
```

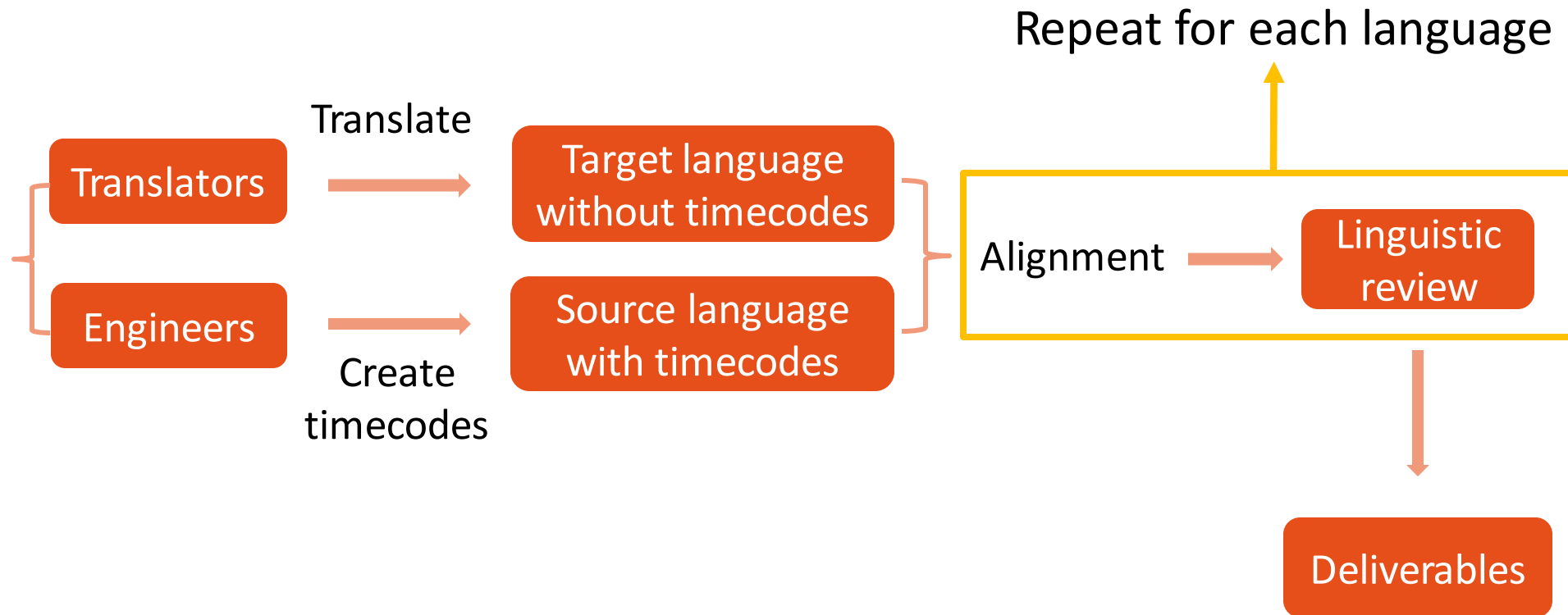
Translation without timecodes



Alignment

Start	End	Source	Target
00:02.391	00:06.228	Have you ever translated a video script into several languages	¿Alguna vez ha traducido un <u>guión</u> de video a varios idiomas y descubrió que cada segmento de la voz en off multilingüe no tiene la misma duración que el original? Por lo tanto, debe hacer coincidir el video y la voz en off manualmente. Eso significa cientos de horas de trabajo aburrido esperando ser hecho. Pero ahora tienes <u>VideoLocalize</u> , la solución perfecta. <u>VideoLocalize</u> puede sincronizar video y voz en off automáticamente con solo unos pocos clics. Como resultado, le ahorra mucho tiempo y dinero. <u>VideoLocalize</u> también proporciona un método de grabación de estilo karaoke, por lo que puede sincronizar tanto el audio como el video sin estirar el video. <u>VideoLocalize</u> no es solo una herramienta de sincronización, sino que también sirve como una plataforma de locutores donde locutores multilingües de todo el mundo ofrecen sus servicios. <u>VideoLocalize</u> también es un entorno de trabajo conjunto en línea para todos los involucrados en un proyecto. Traductores, directores de proyectos, locutores y clientes pueden trabajar en una plataforma para facilitar la comunicación. Fácil localización de videos.
00:06.228	00:09.747	and found out that every segment of	

Translation without timecodes



Translation with timecodes



Transcription with timecodes

```
1
00:00:03,937 --> 00:00:06,106
Security is a full stack exercise:

2
00:00:06,106 --> 00:00:07,957
from the hardware at the base

3
00:00:07,957 --> 00:00:10,777
to the operating
systems that might run upon it,

4
00:00:10,777 --> 00:00:13,747
to the transport protocols
that carry bits back and forth

5
00:00:13,747 --> 00:00:16,566
among different nodes up
through the software running

6
00:00:16,566 --> 00:00:19,936
on those operating systems
and of course, people,
```

Subtitle Rules:

- Minimum duration for a segment is 1 second
- Maximum duration for a segment is 6-7 seconds
- Maximum length per line is 35-48 characters
- Characters per second should be 12-20 CPS
- No more than 2 lines per segment

English	
CPS: 16 Duration: 2.169s 34 characters	1 00:00:03,937 --> 00:00:06,106 Security is a full stack exercise:
CPS: 16 Duration: 1.851s 29 characters	2 00:00:06,106 --> 00:00:07,957 from the hardware at the base
CPS: 17 Duration: 2.820s 16 31 characters	3 00:00:07,957 --> 00:00:10,777 to the operating systems that might run upon it,
CPS: 19 Duration: 2.970s 26 30 characters	4 00:00:10,777 --> 00:00:13,747 to the transport protocols that carry bits back and forth
CPS: 19 Duration: 2.819s 24 28 characters	5 00:00:13,747 --> 00:00:16,566 among different nodes up through the software running
CPS: 15 Duration: 3.370s 26 22 characters	6 00:00:16,566 --> 00:00:19,936 on those operating systems and of course, people,

Spanish – Latin America

CPS: 19 Duration: 2.169s 42 characters	1 00:00:03,937 --> 00:00:06,106 La seguridad es un ejercicio a todo nivel:
CPS: 16 Duration: 1.851s 29 characters	2 00:00:06,106 --> 00:00:07,957 desde el hardware en la base,
CPS: 17 Duration: 2.820s 23 25 characters	3 00:00:07,957 --> 00:00:10,777 los sistemas operativos que se ejecuten sobre él,
CPS: 18 Duration: 2.970s 28 23 characters	4 00:00:10,777 --> 00:00:13,747 los protocolos de transporte que llevan y traen bits
CPS: 19 Duration: 2.819s 28 26 characters	5 00:00:13,747 --> 00:00:16,566 entre distintos nodos hasta el software que se ejecuta
CPS: 17 Duration: 3.370s 28 27 characters	6 00:00:16,566 --> 00:00:19,936 en esos sistemas operativos. Y, claro, también la gente.

Traditional Chinese

CPS: 4 Duration: 2.169s 9 characters	1 00:00:03,937 --> 00:00:06,106 安全性是種全端運用
CPS: 4 Duration: 1.851s 7 characters	2 00:00:06,106 --> 00:00:07,957 從硬體架構開始
CPS: 4 Duration: 2.820s 10 characters	3 00:00:07,957 --> 00:00:10,777 到其中執行的作業軟體
CPS: 3 Duration: 2.970s 9 characters	4 00:00:10,777 --> 00:00:13,747 以及在各個節點之間
CPS: 5 Duration: 2.819s 13 characters	5 00:00:13,747 --> 00:00:16,566 來回傳輸位元的傳輸通訊協定
CPS: 7 Duration: 3.370s 14 8 characters	6 00:00:16,566 --> 00:00:19,936 再到這些作業軟體上運作的軟體 當然還有人事方面

Pros & Cons

	With Timecodes (broken)	Without Timecodes
Pros	<ul style="list-style-type: none">• Delivery as soon as the translation is done• Saves engineering hours	<ul style="list-style-type: none">• Easier for translators to work on• Easier to maintain translation quality and memory
Cons	<ul style="list-style-type: none">• Big challenge for translators• Bad for building Translation Memory• More challenging to translate easily	<ul style="list-style-type: none">• More engineering hours, complicated workflow for multiple languages• Longer delivery time

Types of Transcription

Transcription without timecodes

And sometimes what's important about hearing a message is hearing it more than once because we hear it in unique ways. We're in a new place today, we're in a new season. And so I'm going to give you some strategies and I'm going to share with you kind of where our focus is for our time together.

You're going to learn how to develop a courageous or optimal mindset. So, I want you to touch your head quick. So, an optimal mindset, a courageous mindset, one that is working to your collective advantage. It's like you are at your best.

Then the second area of focus that we're going to cover is we're going to discover a few EQ strengthening techniques and that will support you to navigate each opportunity with mindfulness, confidence, compassion, and enthusiasm. Okay, so mindfulness, confidence, compassion, and enthusiasm, jazz hands.

Okay? All right, so, this is our focus. This is the direction in which we are going to travel today and let's take a quick peak. So, we talked about this at conference, emotional intelligence, EI, emotional intelligence. This is how our ability to read people, to have some self-awareness, to be able to navigate conversations so they're smooth and people leave that conversation feeling like you get me, you understand me, or you shared with me just what I wanted to hear today.

This really resonated with me. When we have this high emotional intelligence, we're able to be agile in our communication. And this is the definition we shared with you, being intelligent about our emotions. So, being intelligent about our emotions and then using emotions to unlock our intelligence. So, being intelligent about our emotions and using our emotions to unlock intelligence.

Transcription with timecodes

```
1
00:00:03,937 --> 00:00:06,106
Security is a full stack exercise:

2
00:00:06,106 --> 00:00:07,957
from the hardware at the base

3
00:00:07,957 --> 00:00:10,777
to the operating
systems that might run upon it,

4
00:00:10,777 --> 00:00:13,747
to the transport protocols
that carry bits back and forth

5
00:00:13,747 --> 00:00:16,566
among different nodes up
through the software running

6
00:00:16,566 --> 00:00:19,936
on those operating systems
and of course, people,
```

Pros & Cons



	With Timecodes	Without Timecodes
Pros	<ul style="list-style-type: none">• Delivery as soon as the translation is done• Saves engineering hours	<ul style="list-style-type: none">• Easier for translators to work on• Easier to maintain translation quality and memory
Cons	<ul style="list-style-type: none">• Big challenge for translators• Bad for building Translation Memory• More challenging to translate easily	<ul style="list-style-type: none">• More engineering hours, complicated workflow for multiple languages• Longer delivery time

Things to consider

- How many languages is it going to be translated into?
- Is building translation memory necessary for future projects?
- What is the requirement for translation quality?
- Is this project urgent?

On-Screen text (OST) – Is the source project file available?

On-Screen Text (OST)


**Facteurs déclenchants
du psoriasis**

PRÉDISPOSITION GÉNÉTIQUE

~~MALADIE PSYCHOSOMATIQUE~~

FACTEURS EXTERNES

- frottements (coudes et genoux)
- maladies infectieuses
- médicaments
- alcool
- tabac
- stress



BIODERMA
LABORATOIRE DERMATOLOGIQUE

Source Project File Types for Videos

Video Editing Software	File Name
Adobe Premiere	Untitled. prproj
Adobe After Effects	Untitled. aep
Camtasia	Untitled. tscproj
Final Cut Pro	Untitled. fcpx
DaVinci Resolve	Untitled. drp

- Comes with a package containing media files, such as raw footage, pictures and audio.

Problems and Solutions – When the source project is not available.

If text cannot be removed or covered.



If the text can be removed or covered.



UI text in the Video – Has the platform been localized?

UI Text

VideoLocalize Pricing Products Dashboard My Account (apps@videolocalize.com)

Welcome to VideoLocalize's Free Subtitle Burning Tool!

Our easy-to-use tool can quickly burn subtitles into your video. To get started, just follow the easy steps below.

Burned-in subtitles, also called open captions, are subtitles that are "burned" into the video itself and cannot be turned off, unlike closed captions, where the subtitles can be turned on or off by the viewer. With open captions, you can also determine the style of the captions – the color and font size – in advance.

Step 1: Upload your video

Upload from computer
 Use Existing Video
 Use YouTube or Direct Link

Choose file(s) OR drop file(s) here

Enter your email address to receive an email with the download link of your completed video, once processing is done.

To receive the video directly, link with Dropbox

apps@videolocalize.com

Start Processing

Step 2: Upload your subtitle file(s)

Choose file(s) OR drop file(s) here

Step 3: Choose subtitle styles

Subtitle Color:

Font: Arial (Unicode)

Subtitle preview

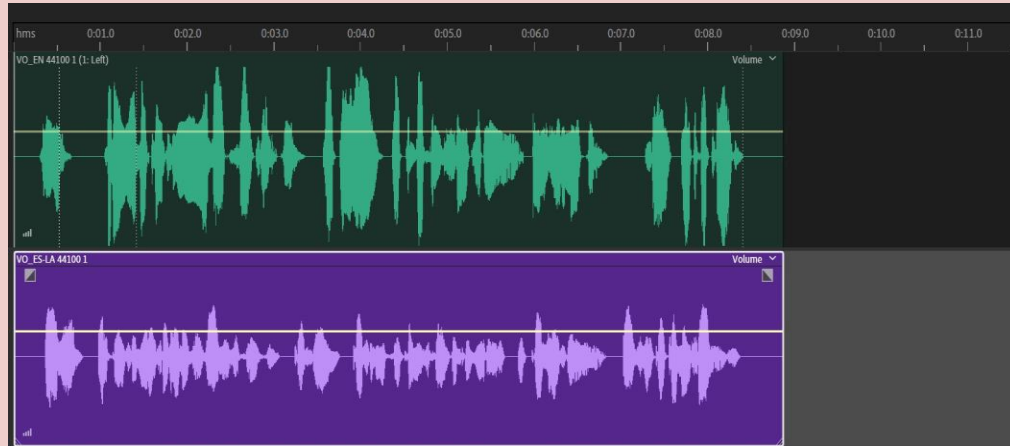
Thanks for using VideoLocalize's free Subtitle Burning Tool

Dubbing Questions

- Whether to dub the video using timed vs untimed recording?
- What type of voiceover to choose: AI machine generated (Text-to-Speech) voices vs human voices?

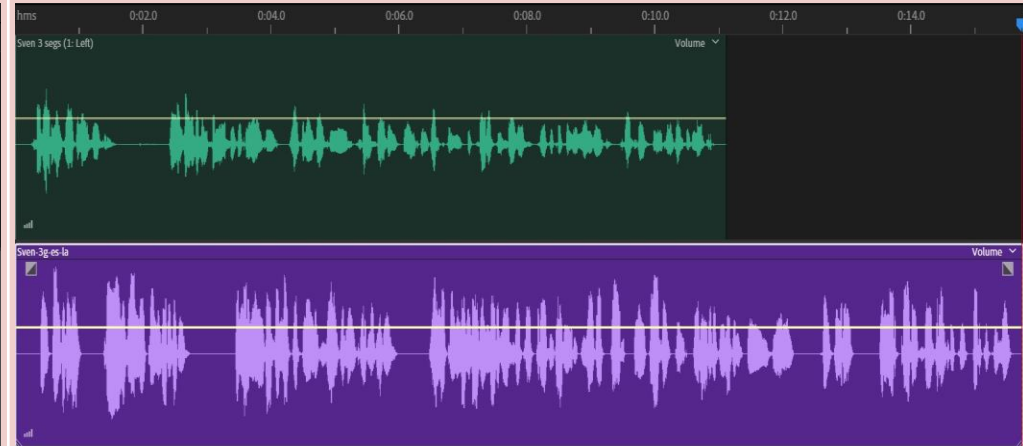
Timed vs. Untimed Recording – which one to choose?

Timed



The voice over recording matches the length of the original recording.

Untimed

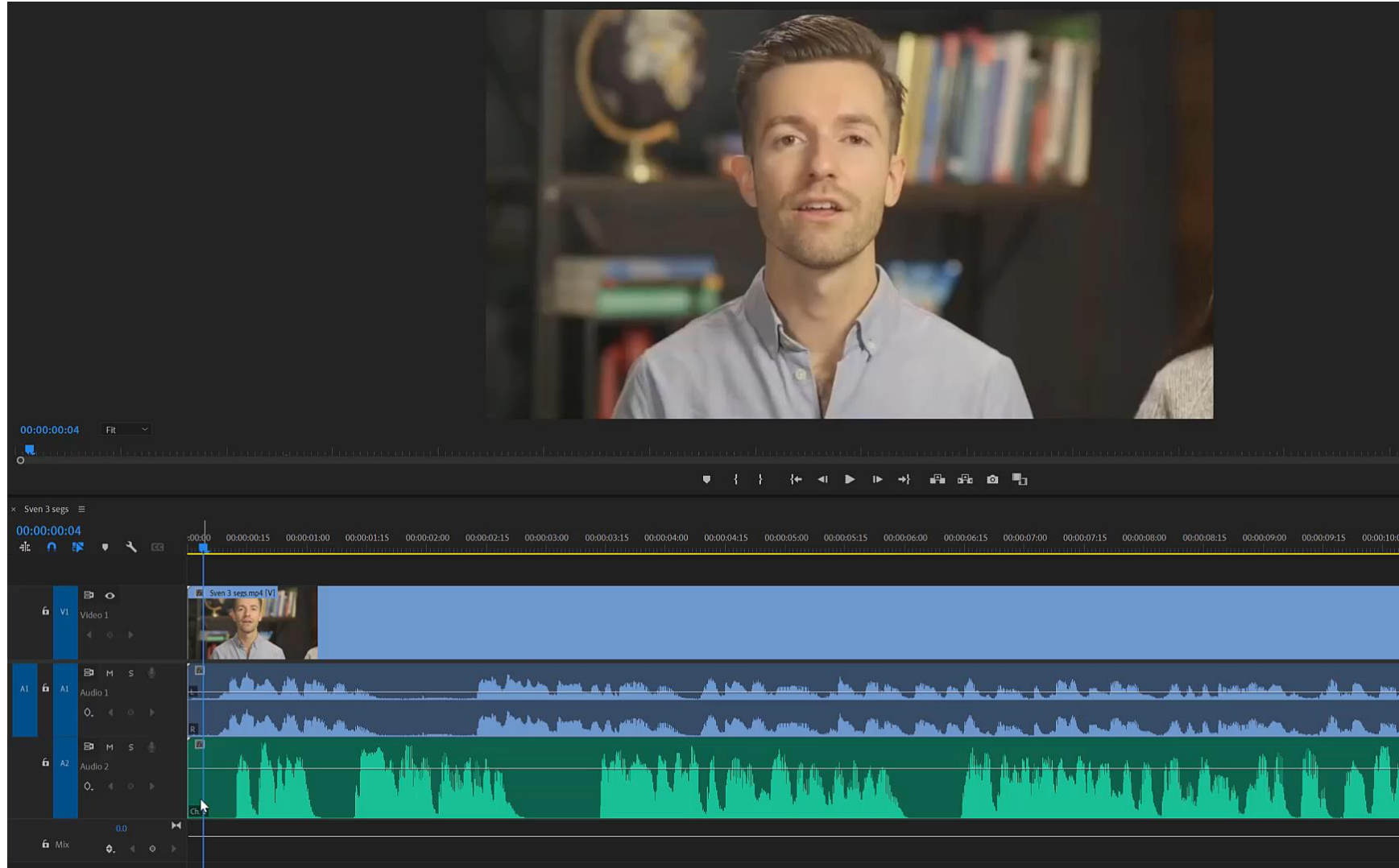


No limitation on the recording.

Pros & Cons

	Timed Recording	Untimed Recording
Pros	<ul style="list-style-type: none">• Accurate, video length stays the same• Simple replacement of the original audio in postproduction	<ul style="list-style-type: none">• Lower cost, shorter studio time• No changes to the voiceover script
Cons	<ul style="list-style-type: none">• Expensive, not all voice talents can do it• Turnaround time is longer• Possible changes to the voiceover script	<ul style="list-style-type: none">• Requires synchronization work• Change of video length

Solutions for Untimed Recordings - Manual



Things to consider

- Does the video length need to stay the same after dubbing?
- What is the budget and turnaround time?

Text-to-Speech (TTS) vs. Human Voiceover

Text-To-Speech (TTS)



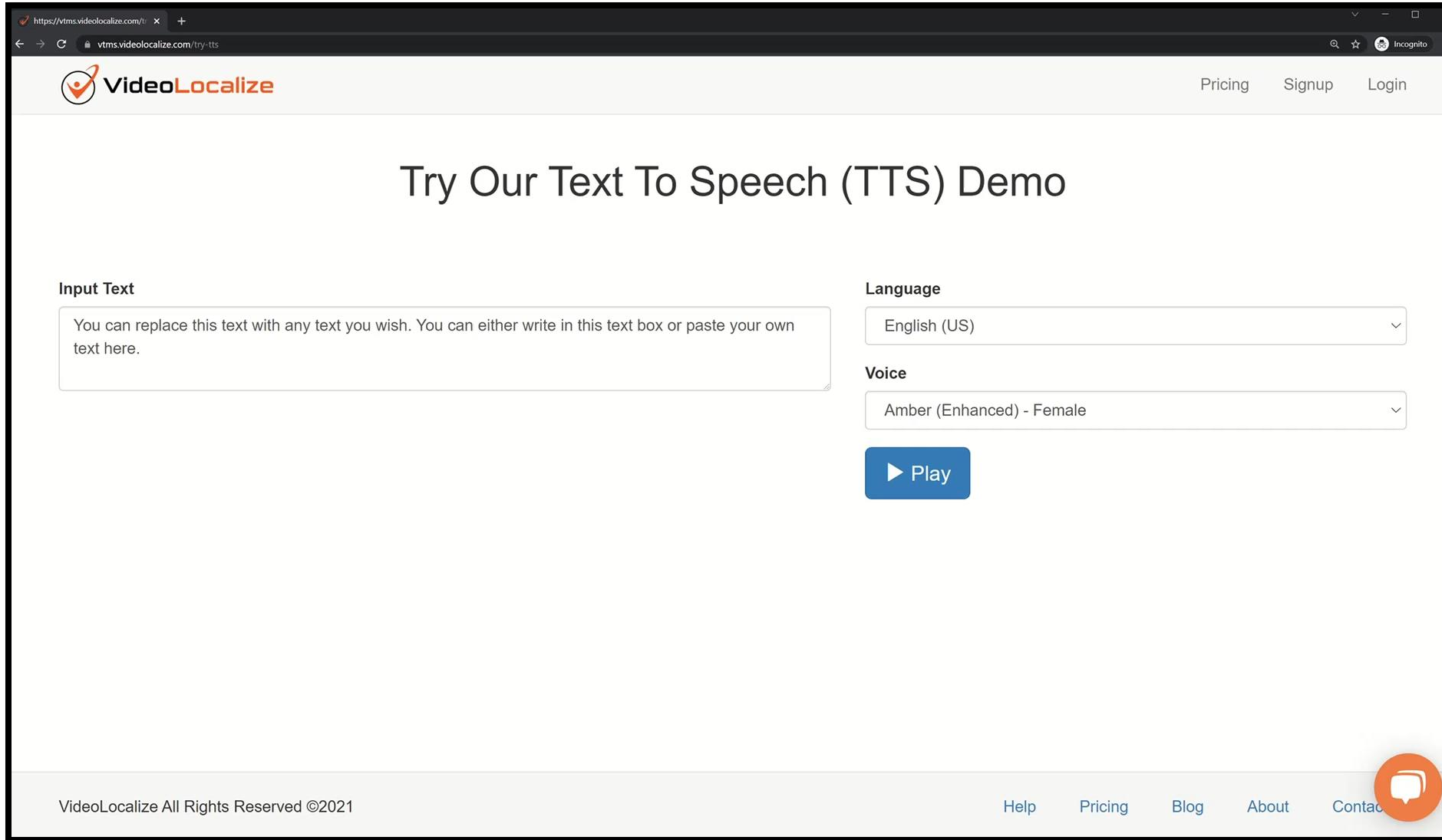
AI computer-generated voice that converts text into spoken words

Human



Voice-over talents

Try TTS Samples in your language: <https://vtms.videolocalize.com/try-tts>



The screenshot shows a web browser window displaying the VideoLocalize TTS demo page. The browser's address bar shows the URL <https://vtms.videolocalize.com/try-tts>. The page features the VideoLocalize logo in the top left corner and navigation links for Pricing, Signup, and Login in the top right. The main heading is "Try Our Text To Speech (TTS) Demo". Below this, there is an "Input Text" section with a text box containing the placeholder text: "You can replace this text with any text you wish. You can either write in this text box or paste your own text here." To the right of the text box are two dropdown menus: "Language" set to "English (US)" and "Voice" set to "Amber (Enhanced) - Female". A blue "Play" button with a play icon is positioned below the dropdowns. The footer contains the text "VideoLocalize All Rights Reserved ©2021" on the left and navigation links for Help, Pricing, Blog, About, and Contact on the right, along with a chat icon.

Text-to-Speech (TTS) vs. Human Voiceover

Text-To-Speech (TTS)





AI computer-generated voice that converts text into spoken words



Human



Voice-over talents

SSML Example

Before	Compatibles con las bases de datos MySQL y SQL.	
After	<pre><speak version="1.0" xmlns="http://www.w3.org/2001/10/synthesis" xml:lang="es-LA"> Compatibles con las bases de datos <phoneme alphabet="ipa" ph="maɪ si kwɔl">MySQL</phoneme> y <phoneme alphabet="ipa" ph="si kwɔl">SQL</phoneme>. </speak></pre>	

Before	All-in-ones u ordenadores de sobremesa.	
After	<pre><speak version="1.0" xmlns="http://www.w3.org/2001/10/synthesis" xml:lang="es-LA"> <phoneme alphabet="ipa" ph="ɔl in wɔnz">All-in-ones</phoneme> u ordenadores de sobremesa. </speak></pre>	

Changing the Language with multilingual voice		
<pre><speak version="1.0" xmlns="http://www.w3.org/2001/10/synthesis" xml:lang="es-LA"> <lang xml:lang="en-US">All-in-ones</lang> u ordenadores de sobremesa. </speak></pre>		

Pros & Cons

	Text-to-Speech	Human
Pros	<ul style="list-style-type: none">• Fast turnaround time• Lower cost• Recordings can be generated instantly if script changes• Can be generated by anyone, no technical background needed	<ul style="list-style-type: none">• Natural sound• All words can be correctly spoken with a pronunciation guide• More voice and language options available
Cons	<ul style="list-style-type: none">• Voice quality varies depending on languages• Limited voice options per language• To adjust pronunciation and prosody requires knowing Speech Synthesis Markup Language (SSML)	<ul style="list-style-type: none">• Longer turnaround time• Higher cost compared to TTS• Any script change requires re-recording• Require voice professionals and studio

Thank you!

Have more questions?
Email us at
teamto@videolocalize.com