

## Audio and Video Localization

Whether you are considering localizing an eLearning course, a video game, or a training program, the audio and video components are going to be central to the project. The following discussion assumes that the English version of the product already exists and that localization takes place as an afterthought – a very pragmatic approach since most localization projects fall in this category. However, the savvy project manager will keep the steps outlined in this document in mind when developing his or her next project if it's likely to be localized.



The screenshot shows a video player interface for a video titled "LMS100". The video content features a man, Tomáš Uoker, speaking. Below the video, there is a poll in German asking "Wem stimmen Sie zu?" (Whom do you agree with?). The poll options are "Eva" and "Karl". The poll text asks the user to choose one answer and click "Fertig" (Finish). The interface also includes a navigation bar at the bottom with buttons for "MENÜ", "BEENDEN", "RESSOURCEN", "MANUSKRIFT", "PAUSE", "WIEDERHOLEN", "ZURÜCK", and "WEITER".

Wem stimmen Sie zu?

**LMS100**

**Томас Уокер**  
Менеджер завода, энергетический центр в Ларедо, Техас

Eva und Karl sind unterschiedlicher Ansicht darüber, welche Rolle der Verhaltenskodex bei ihrer Arbeit spielt. Wessen Meinung teilen Sie?

Wählen Sie **eine** besonders gute Antwort, und klicken Sie auf **Fertig**.

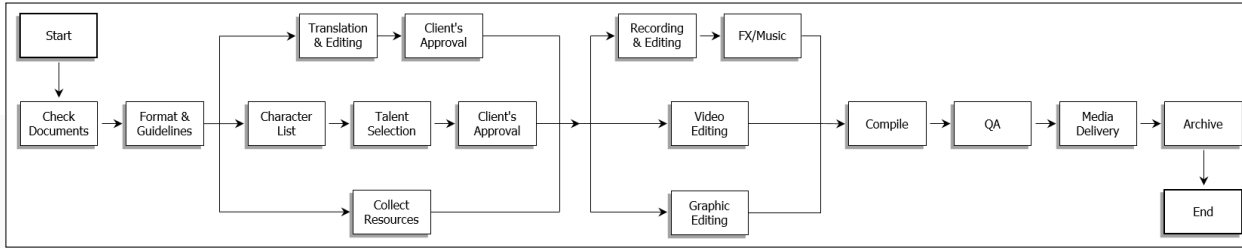
- Eva
- Karl
- Beide
- Keiner

Fertig Neu anfangen

MENÜ BEENDEN RESSOURCEN MANUSKRIFT PAUSE WIEDERHOLEN ZURÜCK WEITER

### Omni Intercommunications, Inc.

2825 Wilcrest - Suite 400  
Houston, Texas 77042  
1-800-777-2304  
info@omni-inter.com



*Audio-Video Localization Process*

When the project involves audio only, the process is fairly simple. Conversely, a complex eLearning project involves many steps besides the audio-video content. In this article, we will consider the localization of a training video which should give a good starting point when considering localization.

Creating a video and localizing a video are fundamentally different. In the first instance (creation), the audio is recorded first, then the video is cut to the audio. In other words, the audio is governing the timing. But in the second instance (localization), the visuals are governing the timing, and the audio has to conform to the video. As we will see, this is not a trivial difference.

### Start with Accurate Documents

The first step of the process is to make sure the project documents match the product to be localized. This involves making sure the script is as-recorded and the screen text is accurate. In most instances, scripts and storyboards are fine tuned during production, but last-minute changes are often left out. Having an inaccurate script can create havoc in the recording studio and translate into cost overruns.

### Format the Documents Correctly

Next, the documents need to be formatted to allow for efficient translation. A Microsoft Word table is typically used for this purpose. Showing the on-screen text in parallel with the narration makes it easier to ensure consistency between the two. An important aspect is to flag what needs to be translated and what needs to be left in English so as to facilitate the work of the translators and production team. As shown below, a Comments column is often required to relay important information to all the parties involved in the production (translators, narrators, reviewers, etc.).

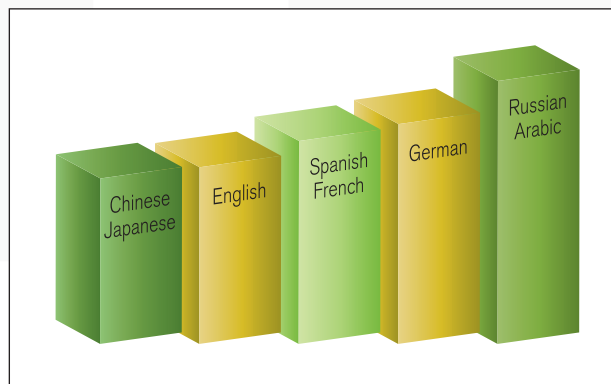
Time Stamp	Onscreen Text - English	Onscreen Text - French	English Voiceover	Voiceover - French	Comments
:06	401(k)		Is a 401(k) really all it's cracked up to be?  You bet!		Pronounced : 4-oh-1-K
:10			A 401(k) is one of the best ways to save for your future and make your nest egg grow faster.		
:17			But it all starts with you setting aside money from your paycheck.		
:19	It all starts with you.		It doesn't have to be a lot.		
:25	Start with as little as 1%		You can start with as little as 1% of your pay.		

### Watch the Timing

When using a table format, the script is divided in cells, each one covering a visual unit. Consequently, the content in these cells is usually of varied lengths, following the cadence of the video. This table format helps the reviewer/editor with the timing of the translation, an aspect which is often overlooked. As a rule of thumb, the source and target language cells should be of similar length, keeping in mind that the shorter the segment, the closer the source and translation have to be. While too much text is going to force the narrator to speak too fast, too little text is going to leave dead air in the final product. In order to avoid these two undesirable situations, the editor needs to review the translation while viewing the video in order to adjust the timing. Ideally, the final adjustments required should be minimal during recording.

Compared to most other languages, English is quite concise. It is not unusual that a translated text be 10-20% longer than the English original, with Russian and Arabic often much longer.

Depending on the budget and the nature of the program, the video may be extended to accommodate the expansion of the text, but most people prefer to edit the translation to avoid additional video work. If the program involves stills like charts and diagrams, editing is minimal and therefore an inexpensive alternative to content editing, assuming the overall length of the program is not an issue.



*Audio Expansion*

### Provide Guidelines

Pronunciation guidelines should be added while preparing the script along with indications of which terms need to remain in English or terms that may have an established translation. It is good practice to review the video for delivery speed, dead air, still graphics, etc. to gauge flexibility with text expansion and mark the script accordingly in the Comments column. A little pre-production work goes a long way in the overall cost and quality of the finished product.

### See the Program from the End User's Point of View

Although this is not specific to audio-video production, it is important to localize the characters' names, the cities, the currencies, the distances and temperatures, and make sure the idioms are correctly adapted. Ideally, a localized product should never be perceived as a translation but rather as a product written specifically for the end user. Although it is unlikely that this outcome be totally achieved, removing as many barriers as possible facilitates the acceptance and understanding of the message. Just like a distance of 220 meters does not "speak" to an American audience, a "two football field" long warehouse gives no clue to its German counterpart.



### Split the Process

At this point, the process splits between a language side and a production side. On the language side, the documents will go through translation, review and editing, and finally, client review and approval. Depending on the project, the translation may go back and forth between the translation team and the client's approvers until all questions have been resolved and the translated documents are approved and ready for production.

### Get the End User Involved

The importance of the end user's review and approval cannot be over-emphasized. Note that 'client' and 'end user' are often two different entities, the client being the party ordering the work, while the end user is the party using the work. The end user or the client's local office in France or China, for instance, will know the correct terminology and whether the processes, regulations, laws, etc. referred to in the training are applicable. Conversely, country-specific procedures may have been omitted. In addition, involving the end user in the development of the training program will give them a sense of ownership and increase acceptance of the training. The reviewer needs to be made aware of the above-mentioned timing constraints and asked to comment on the pronunciation guidelines, especially on acronyms. Obviously, this review is not intended to shift the responsibility of the translation to the client; it's there to ensure that a final level of polishing has been achieved. As such, the work required should be minimal.

While the translation is underway, much work has to be performed on the production side.

### Select the Right Talents

The first step is to select the right voice talents to be used for recording. While corporate videos typically require a single voice, training programs often times require several. In each case, it is again important that the end user be involved in the selection of the voices. What may sound perfect to a non-native speaker may in fact be unacceptable, since dialects and accents are of prime importance to the success of the final product. Just like a safety program to be shown in Texas would not use a British actor, one to be shown in Angola should not use a Brazilian narrator.

Britain and America are two nations divided by a common language.

Be mindful of the local customs as well. While it is perfectly acceptable to use female voices for almost any program in the US or Europe, other parts of the world may not be so receptive. Again, the end user's input is important.

Although translation cost is often the most important item in the overall price of localization, the cost of narrators may become excessive in the case of role plays or testimonials. In order to keep the number of voices to a minimum, it is necessary to create a character list and determine the number of voices needed. This depends on the dubbing style selected (see [Annex A](#)). With an interview style, two to three voices are usually enough, i.e., one for the male characters, one for the female, and one for the narrator. With lip-synch, since the idea is to give the impression that the person on the screen actually speaks the language, each character needs to be assigned a distinctive voice. However, it is often enough to make sure that the same foreign language talent is not assigned to two characters appearing on the screen at the same time.

Role Play	Character	Sex/Age	Talent
Vignette 1	Mary	Female/30	Natacha S.
	John	Male/50	Mikhail B.
Vignette 2	Mary	Female/30	Natacha S.
	Tania	Female/30	Vita F.
Vignette 3	Roger	Male/20	Lev T.
	Chris	Male/40	Mikhail B.
Vignette 4	Roger	Male/20	Lev T.
	Kimberly	Female/50	Vita F.
Total	6		4

Using a list of character interactions, the producer assigns each character to a talent while minimizing their total number. Keep in mind though that in certain cases, even though two characters may never appear together, it is necessary to use additional talents, for instance if there is a child in an otherwise all adult production. Conversely, if the production is very long, adding talents may not increase the price substantially if each talent is used for longer than the time covered by his or her minimum charge. As with any production, an experienced producer should be able to balance cost vs. overall quality.

Select voice talents based on locale, gender, age, etc. Choose people who are fluent in the language the way it's spoken in

the target market since accent, speed, and inflection can easily be misjudged. Use professional narrators. They are more expensive than an employee or a student, but it will make a huge difference in the finished product and the cost of the recording studio. A professional is also more likely to be available for pickups or edits down the line and is trained to have a reproducible voice. All selected talents should be approved before actual production begins.

### Prepare for Production

Now that the narrators have been selected, resources need to be collected in their native format if possible to simplify the localization. A split audio is preferable to a mixed one since the ambient sounds, also known as FX, can be re-inserted easily in the foreign language version (a production without background noises can be very dull!). If appropriate, the music track can be reused (subject to licensing). Layered, text-editable Photoshop or video files facilitate graphic editing and result in a cleaner product. If not available, boxes need to be used to block the source text, resulting in increased production costs and inferior esthetics if the background is not uniform.



### Recording Best Practices

When it's time to record, make sure the studio records in synch with the video, and not separate audio tracks with time code/segment duration as its only guideline. This will ensure the tone is right for each scene and that the delivery is not artificially compressed, for instance when there is a pause following the English narration.

This does not apply to most eLearning products since the audio usually controls the delivery, i.e., one audio file needs to end for the next screen to load (some eLearning developers use markers in a single audio file, leading to the same result but with slightly more editing work since the markers need to be inserted in each language).

The audio technician needs to be familiar with the language. Often what might sound like a lip noise at the end of a word or sentence is an actual part of the speech. An alternative is to use a bilingual person to assist the editor, but this complicates the editing process with a resulting negative effect on cost.

It is also a good practice to keep track of the equipment settings for each recording. Should pick-ups be required at a later date, these settings will be invaluable in matching the recordings.

Usually, the talents are recorded separately in a linear fashion, each one being assigned to a given audio track in order to facilitate editing. Once the main character has been recorded, the following characters can interact with him/her since the audio technician will pre-roll the foreign language instead of the English one.

It is not unusual for minor edits to be done to the script at the recording stage but this should be kept to the bare minimum for two main reasons. First, these edits lengthen the production time and consequently, the production costs. Second and more importantly, depending on the material, what may look like a minor edit to the talent or producer may be a sensitive issue for the client, making the edit unacceptable. It is therefore good practice to stick to the approved script.

Once the narration has been completed, the FX and music are added.

Always work with high quality audio and video. Do not record in .mp3 for instance, but rather in .wav or .avi format. Do not compress until everything has been approved. This will facilitate edits down the line and allow for distribution across multiple media (DVD and online for instance).

### Insert the Graphics into the Video

In the last step of the process, the video technician either localizes the graphics directly on the video track, or replaces graphics with localized versions prepared separately. The method selected depends on several factors such as the complexity of the graphics and the availability of editable video tracks.



### Compile the Various Resources

If the final product is a complex eLearning program involving online quizzes or a complex user interface, this phase usually requires extensive engineering which is beyond the scope of this article. However, for videos and simple eLearning modules, the final production will mimic that of the source material and does not need extra explanation.

### QA the Final Product

Quality assurance (QA) should be the final step of the process. Ideally, this QA should be performed by a native speaker who has never seen the English program to make sure the localized version makes sense on its own. As with the English version, inconsistency between text on the screen and narration, typos, missing words or punctuation, and truncated audio files are among the items that will most likely surface and require correction. Once completed, the product is ready for final compilation and delivery under the format(s) required for distribution.

## Omni Intercommunications

Founded in 1978 in Houston, Texas, Omni Intercommunications is a foreign language communications company that has built a solid reputation of helping multinational organizations achieve greater success through effective translation and localization. We can work in any language and culture, in any part of the world our clients need to reach. We produce turnkey products in our in-house audio/video and graphic studios. Our team of professional language and culture experts, skilled graphic designers, programmers, and project managers, are dedicated to providing world-class service at every phase of the production process, regardless of project, language, or magnitude. Whether your company needs translation services for print, audio/video, eLearning, or software localization, you'll find that the Omni team can leverage its strengths, skills, and expertise to deliver solutions that translate into real value for your business. We don't just translate – we produce.

## Annex A - Video Localization

There are essentially two methods for creating a foreign language video depending on whether the audio track is replaced (dubbing) or not (subtitling/closed captioning). Although both methods can convey the original message to a foreign audience, they are fundamentally different. The cost savings of subtitling must be carefully weighed against the much greater effectiveness of dubbing. Since each method has its pros and cons, make sure you understand the differences to determine what works best for your budget and application.

### **Dubbing with Lip-Synch (click for a sample)**

For training videos with on-camera speakers, the most efficient technique is lip-synching, where the person on screen appears to be speaking the foreign language. This requires extensive adaptation of the original text in order to match the lips. Most Hollywood movies that have been dubbed employ the lip-synching method.

### **Dubbing with Interview Style (click for a sample)**

Also known as U.N. style, interview style makes no attempt to follow the lips, but only the general rhythm, expressions, and pauses of the speaker. The English is generally kept low in the background to ensure a better sound experience for the audience. This is the perfect approach for testimonials where it is known that the people on screen do not speak the language, or at least did not use it at the time of the recording.

### **Dubbing with Localized Voice-Over (click for a sample)**

If the video does not contain anyone on camera, the voice-over track is simply replaced altogether. Localized voice-over is the easiest to produce since it typically involves a single narrator and time constraints can be handled by an experienced translator. Most corporate and travel videos fall in this category.

### **Subtitling or Closed Captioning (click for a sample)**

The least expensive video localization format, subtitling or closed captioning, allows the viewer to read the translated text on screen while listening to the original audio. The text can either be permanently burned-in as part of the video (subtitling) or separated (with closed captioning, the translation is encoded and enabled at the viewer's discretion, much like on a DVD). Subtitling should be considered carefully for training videos since the attention of the viewer is on the subtitles rather than on the behavior or procedure demonstrated. It is however a viable alternative for testimonials.