White Paper

How to Improve the Efficiency and Quality of Audio Dubbing in File-Based Media Workflows

Niraj
Interra Systems, Inc.

1. INTRODUCTION

In recent years, overcoming language barriers has become exceedingly important for content creators. The advent of digital television, along with its increasing adoption around the globe and rapid boom in mass media delivery over the internet, has made this requirement even more prominent today. There is an increasing need to maximize the outreach of content across geographies, even to the non-native speakers that are incapable of understanding native language.

This can be achieved either through the use of captions and subtitles or audio dubbing to regional languages. Subtitles is an easier and more affordable option, but with many drawbacks. For instance, subtitles are limited in time and space. Moreover, with subtitles, it can be difficult to effectively deliver long dialogue scenes within limited screen times. Furthermore, subtitles are distracting in nature. Since viewers' focus is on the text on the bottom of screen, many details in the scene can be lost. Additionally, subtitles only convey the meaning of what is being said; they do not capture the emotions that are being conveyed in the dialogue. Overall, subtitles are primarily useful for literate viewers with fast reading skills.

Due to the limitations of subtitles, audio dubbing to regional languages is a preferred option in many cases. Dubbing is a process of translating a foreign language program into the audience's native language. Dubbed tracks are created by adding language-specific content to the original audio. While audio dubbing used to be costly a few years back, in recent years, thanks to advanced low-cost dubbing applications and more affordable dubbing artists, it has become more reasonable. It is increasingly being adopted by content manufacturers around the globe. Digital audio files are being dubbed and aired on TV or hosted on OTT servers like YouTube. A great example of audio dubbing is when the same television program is aired in different languages, which can be selected on the set-top box menu.

Given the popularity of audio dubbing, content creators and broadcasters today have a greater responsibility to ensure proper quality of dubbed audio tracks and synchronization with the video or original audio track.

2. AUDIO DUBBING CHALLENGES

When content is prepared in different languages, it is then multiplexed with the corresponding video and stored as separate tracks in the digital files. Typically, each track provides audio in a different language. Ensuring the audio with the right language is on the correct track is a cumbersome manual task. It is particularly difficult for one person, as there can be multiple language audio tracks in a single file. One person would typically not be fluent in all of the languages.

The audio dubbing process is not trivial. Dubbing audio in multiple languages is accomplished by separately dubbing artists with proficiency in each of the respective languages. For each language, there is a dubbing translator and an actor involved. The responsibility of the translator is to create a language transcript that is converted to audio by a dubbing actor.

A major challenge during the dubbing process is synchronization between audio and video throughout the stream. Even a minor lead or lag of 500 milliseconds for a small portion of content is not acceptable if content creators want to ensure a high-quality viewing experience. It is the responsibility of the dubbing translator to make words match the visual that appears onscreen. The length of sentences should be small; otherwise, the actor would not be able to pronounce them within the required duration.

Since the same idea is conveyed in different languages in sentences of remarkably different durations, keeping synchronization across master and dubbed tracks can be very challenging. There is a high probability that sync issues will show up in the dubbed tracks, starting at transcription and recording stages. Issues might also arise in later

stages of the workflow, with the loss of one or more frames in any dubbed track. There should be a way to check for loss of sync between the dubbed track and master track. It is extremely difficult to do this manually, as it requires listening to the entire piece of content and labeling lead or lag for each portion separately.

It is also possible that the original audio track is modified after the dubbing stage. This occurs when a portion of the audio track has been removed or added. In such cases, a mismatch of duration happens between the original and dubbed tracks. It is challenging to check for metadata-related mismatch between the original and dubbed tracks, particularly when the number of dubbed tracks is very high within the file.

To make dubbing workflows more efficient, these challenges need to be tackled head-on. Using affordable QC solutions, content creators can streamline the audio dubbing process.

3. QUALITY CHECKING FOR DUBBING WORKFLOWS

With the advances in digital media technology and machine learning (ML) in recent years, it has become possible to automate verification of language and checking of synchronization loss in dubbed audio tracks using media QC tools. A QC solution for dubbing workflows should have following key capabilities:

Verifies dubbing packages with complex structure: The QC solution should be able to verify complex dubbing packages. Many times, a dubbing package is not a single file but consists of multiple files. For example, the dubbing package may include MXF files and multiple .wav files. The MXF file contains video tracks along with original audio tracks and .wav files corresponding to multiple dubbed audio tracks. Sometimes .wav files represent individual channels of a 5.1 audio track. Another package may have multiple audio tracks or channels encapsulated in container formats instead of .wav files. The QC solution should be able to handle all of these package variations and verify the multiple dubbed audio tracks properly.

Metadata verification: Metadata verification is an important part of any media QC solution. It should be able to verify the structure of the package (i.e., the number of audio tracks or the channel configuration of multiple dubbed tracks). It should also be able to verify and compare the duration of the original audio track with that of the dubbed audio tracks.

Language detection and verification: Language identification should be an integral part of any QC solution for audio as broadcasters and post-production houses need to ensure that each track has the right language meant for it when transmitting content targeted for global audiences. With machines getting more intelligent and computing power becoming more affordable, it is now possible to automate language detection of audio with great accuracy. This requires availability of a few hours of audio content in the target languages, which are then used to train the ML models. Once trained, the models can predict the language of audio sequences presented to them. The predicted language can be verified against the metadata to ensure that the track has the correct language.

Synchronization between video and dubbed tracks: It is difficult to check synchronization loss between video and any audio track. There is no way to map the correct audio sequence for any video screen with dialogue. But, many times video contains black frames and color bars deliberately put in the program for various requirements, including synchronization. The corresponding audio sequence for black frames and color bars are silence and test tone, respectively. The QC solution should be able to verify that black frames in video tracks appear along with silence at the same time and color bars appear along with test tone in dubbed audio tracks. This enables content creators to check for synchronization loss between video and audio tracks.

Synchronization between original and dubbed tracks: Checking for synchronization loss between the original and dubbed audio tracks is not a minor issue. The reason is that audio data in both master and dubbed tracks are completely

different. But in most cases, both the master track and dubbed track have common background music or effects. Background music and effects can be separated from the audio track using various mechanisms like band pass filter. Localized correlation between background beds of dubbed audio tracks and the original audio track needs to be performed to find any loss of synchronization. The challenge is ensuring proper separation of background bed from audio tracks. One way this can be achieved is by comparing loudness curves. The loudness curves between the original and dubbed track follow same pattern when there is no sync loss. This feature would be quite useful in QC solution, allowing content creators to check for any mismatch between loudness values of original and dubbed tracks.

4. STREAMLINING THE DUBBED AUDIO PROCESS IN FILE-BASED WORKFLOWS USING BATON

Interra Systems' BATON® is the leading enterprise-class QC solution used by global telcos, broadcasters, post-production houses, IPTV and archiving companies working with file-based content. With BATON, operators can efficiently, accurately, and cost-effectively tackle the majority of dubbing workflow requirements.

An advanced language detection feature in BATON uses advanced ML algorithms to detect language in any audio track with more than 90 percent accuracy. This feature is highly customizable. The detected language in audio tracks can be verified against the expected language defined by users. There is also an option to verify the detected language against metadata from the audio track.

BATON supports all the variations of media file packaging used in the industry. Even if the dubbed audio track is an external .wav file or wrapped in another container format, BATON is able to support it. There is a highly configurable option to merge or distribute file-based packaged audio tracks in any way.

BATON supports metadata-related comparisons between all audio tracks in the file.

Additionally, BATON supports synchronization checks between audio tracks by comparing the background bed of master audio tracks with that of dubbed tracks. The limitation for this feature is that audio tracks should share the same background music and effects.

BATON supports comparing loudness values between audio tracks. The loudness value is calculated using standard CALM or ITU algorithms. Program loudness is the integrated loudness of the whole track, and the loudness range gives the statistical measure of loudness variation within any track. Using BATON, one can easily calculate and compare the values between audio tracks.

BATON checks for synchronization of video tracks with audio tracks. This highly configurable feature ensures that black frames in video include silence in the audio track. Similarly, it is possible to check that color bars come in parallel with test tone. Users can configure this capability in such a way that an error is reported if color bars come with or without test tone for a given amount of time.

Beyond these features, BATON supports exhaustive metadata and audio quality checks for all audio tracks for nearly all industry audio formats.

5. CONCLUSION

Audio dubbing has become indispensable for content creators around the globe. Because of the sheer volume of content that is being dealt with, there is a necessity to automate QC for dubbing workflows. With the latest ML technologies and advanced audio quality algorithms, it is now possible to QC dubbed packages with higher accuracy than ever before. New AI-driven technology can be used to effectively spot issues with reduced manual supervision. Thus, by deploying AI-based QC solutions like BATON, content creators can introduce efficiency and quality to their dubbing workflows.