

# The Future is Automated

The landscape ahead for M+E looks uncertain. Automation can ease the journey.

## WORKFLOWS AND THE CLOUD

From cloud-based productions to unprecedented storage demands, the supply chain has never been more challenging

## SMART CONTENT

The questions around AI, metadata, and analytics are endless. But we do have many of the answers already

## SECURITY SOLUTIONS:

Cybersecurity and content protection in M&E is always on defense. Can automation can change that?

# 23,01

# A VOICE FOR AUTOMATION

Not every film is a blockbuster.  
And not every film needs human dub work

**ABSTRACT:** Dubbing is the most labor and cost intensive type of media localization, aiming to reproduce the original experience for the target language audience. Automatic video dubbing aims to automatically revoice videos to make them more easily accessible to audiences in other languages at just a fraction of the time and cost.

**By Volker Steinbiss, Managing Director,  
AppTek GmbH**

It is no secret dubbing is the most labor and cost intensive type of media localization, aiming to reproduce for the target audience an experience identical to that of the source language audience.

The dubbing workflow is a complex one involving first a translation of the script in the target language and then its adaptation to fit the time constraints of the video at hand and at the same time achieve synchronicity with the lip movements of the actors on the screen. This script is then used in a recording session with one or multiple actors, depending on the requirements of the video in question, to generate speech that reproduces the emotions of the source speakers, coherent with the body language of the actors on the screen while matching their lip movements.

Automatic video dubbing aims to automatically revoice videos to make them more easily accessible to audiences in other languages at just a fraction of the time and cost. Coming from Germany, traditionally a dubbing country, I understand studio quality dubbing is akin to film production and the acting performances that go with it, and I admit I do enjoy such dubbed content for my favorite blockbusters. Yet not everything we watch online

*NOT EVERYTHING WE WATCH online is a blockbuster. The long tail of content is where I believe automatic video dubbing can make a difference.*

is a blockbuster – in fact very little of the video content people watch online is. It is the latter use cases, the long tail of content that is available and possibly not localized in as many languages as it could, where I believe automatic video dubbing can make a difference.

AI systems today have superhuman abilities in many ways. With regards to language, no human speaks 50 languages, yet computers do. They do so less than perfectly, but when it comes to, say, broadcasting news in 50 languages simultaneously with a negligible delay from the original broadcast so people around the world can get the gist of what is being discussed in almost real time, this is something that a machine is ideal for. Especially when the decision is to localize such a piece of content at a last minute's notice, or where one's budget does not allow for the costs involved in hiring highly skilled simultaneous interpreters in all the languages in question. There is other ephemeral content for which the same applies, such as some types of user generated content, reality TV and so on.

AppTek's fully automatic dubbing pipeline provides an efficient and scalable solution for such content. It consists of a cascaded pipeline of automatic speech recognition (ASR), machine translation (MT) and text-to-speech (TTS) technologies, that are enhanced by features such as automatic speaker grouping and feature extraction, isochronous machine translation that considers the length of source utterances, and speaker-adaptive TTS that mimics the characteristics of the voices of the source speakers.


The results are further enhanced when the ASR and MT components of the pipeline are adapted beforehand to the domain of the video content and the lexicon is updated. When a fully automatic solution is your only choice, due to budgetary or other reasons, it is important to go for the highest quality possible that you can achieve with machines alone, to minimize the number of errors in the final content. In other words, if you have the data, use it so the system learns from it!

When the content requires an even higher level of accuracy, such as media archives and informational videos,

or less-than-premium entertainment content, telenovelas, documentaries or the FAST streaming channels that are the focus of everyone's attention of late, a post-editing workflow is recommended to close the quality gap. In such a scenario, a professional is involved to provide the necessary corrections to the ASR or MT output, or an existing, manually created script file can be used as an alternative. The resulting TTS output can also be post-edited for more natural, emotional flowing speech via an interface that allows users to change the pitch, intonation, speed, and emotion of utterances at the sentence, word or even phoneme level, and which is also available via API for interested customers.

As with all new technologies that reach a satisfactory level of technical quality, they beg the question on how to use them until such time when they become perfect and can replace any manual workflows that existed prior to their development. At present time our automatic dubbing technology is best suited for content that requires a fast turnaround time and has a limited budget. But as the technology continues to evolve, it has the potential to become a viable alternative to manual dubbing workflows. Such development requires partnerships with professionals in the field to ensure the industry's quality standards are consistently met during this journey.

At AppTek, we pride ourselves in helping companies understand where they can make a dent in the technology that would have a significant impact in their business or in the lives of the professionals who do the actual work. This is why we are offering our technology in a modular manner, where users can choose to post-edit any or all the technology outputs in the pipeline, or simply use existing transcripts or translations to achieve the quality standards of their sector. Our goal is to develop user-friendly AI that works — and to do so with partnerships in the sector.

Learn more about AppTek's automatic dubbing by scheduling a consultation today ([info@apptek.com](mailto:info@apptek.com)) and give us your feedback on what is most important to you so we can make sure to take it into account in our roadmap. 



*Dr. Volker Steinbiss is the managing director at AppTek GmbH and staff member at RWTH Aachen University. He holds a PhD in mathematics and worked on speech recognition at Philips Research in the late 1980s, before his interests broadened into more fields of human language technology, such as speech translation, synthetic speech, and natural language processing. [vsteinbiss@apptek.com](mailto:vsteinbiss@apptek.com) @AppTek\_McLean*