

Scale Up Multilingualism in Health Emergency Learning: Developing an Automated Transcription and Translation Tool

Heini UTUNEN^{a,1}, Thomas STAUBITZ^b, Richelle GEORGE^a, Yu (Ursula) ZHAO^c, Sebastian SERTH^b and Anna TOKAR^a

^a*Health Emergencies Programme, World Health Organization, Geneva, Switzerland*

^b*Hasso Plattner Institute, University of Potsdam, Potsdam, Germany*

^c*Digital health and innovation, World Health Organization, Geneva, Switzerland*

Abstract. World Health Organization's (WHO) emergency learning platform OpenWHO provided by Hasso Plattner Institut (HPI) delivered online learning in real-time and in multiple languages during the COVID-19 pandemic. The challenge was to move from manual transcription and translation to automated to increase the speed and quantity of materials and languages available. TransPipe tool was introduced to facilitate this task. We describe the TransPipe development, analyze its functioning and report key results achieved. TransPipe successfully connects existing services and provides a suitable workflow to create and maintain video subtitles in different languages. By the end of 2022, the tool transcribed nearly 4,700 minutes of video content and translated 1,050,700 characters of video subtitles. Automated transcription and translation have enormous potential as a public health learning tool, allowing the near-simultaneous availability of video subtitles on OpenWHO in many languages, thus improving the usability of the learning materials in multiple languages for wider audiences.

Keywords. digital health, artificial intelligence (AI), online learning, health emergencies.

1. Introduction

OpenWHO is the WHO's open source learning platform in health emergencies, providing unlimited access to health professionals and anybody in need of epidemic and pandemic learning. It is a platform provided and adjusted for the purpose by HPI. The Covid-19 pandemic exponentially increased registered users from 160 000 enrolments in January 2020 to more than 7.5 million currently. The platform provides courses on pandemic and epidemic topics in self-paced modalities with videos, quizzes and reference materials.

During the COVID-19 pandemic, OpenWHO needed to deliver lifesaving training materials to frontline health workers in a timely manner and in their own language.

¹ Corresponding author: Heini Utunen, WHO Health Emergencies Programme, World Health Organization. Avenue Appia 20, Geneva, Switzerland; Geneva, Switzerland; E-mail: utunenh@who.int.

However, quick and reliable translation can be difficult and expensive, often requiring manual and time-intensive work [1]. First, there was a lack of automated systems in place, causing repetitive transcription and translation. Second, there was limited translation capacity due to competing tasks. Moreover, results could be undermined by the complexity of the materials [2]. Even when automated transcription and translation services were utilized, the processes of getting videos to these services, retrieving the results, providing quality assurance (QA), and adding the transcripts and translations as subtitles to the videos was a completely manual process. Several steps in this process were highly error-prone. For example, the subtitle file format in use must be machine readable for display with the video. Particularly, manual editing in word processors or sending the files back and forth per mail often resulted in broken subtitle files or wrong character encodings, which then caused errors with the display of the affected videos. Incremental innovation was therefore urgently needed to tackle the above challenges and to mitigate potential risks of failure during the emergency scaling up. Between April 2020 and September 2021 OpenWHO and Hasso Plattner Institut (HPI) created, tested, and finally established TransPipe, a user interface that facilitates the automated transcription and translation solution with possibility of human interference for quality control. TransPipe itself is not an artificial intelligence tool. It, basically, connects the video provider, the learning platform and several transcription and translation tools, which use artificial intelligence to provide their services. The name TransPipe is short for “Transcription and Translation Pipeline”. Its purpose is to fill the bits in between the existing tools and to provide a smooth and streamlined process that eliminates the pitfalls of the previous manual workflow. On top of connecting the existing tools, TransPipe adds a quality assurance workflow that allows human editors, translators, and quality assurance managers to be assigned to a certain course, video, or language. Finally, it visualizes the current state of the process of a certain course, video, or language in a course (see Figure 1).

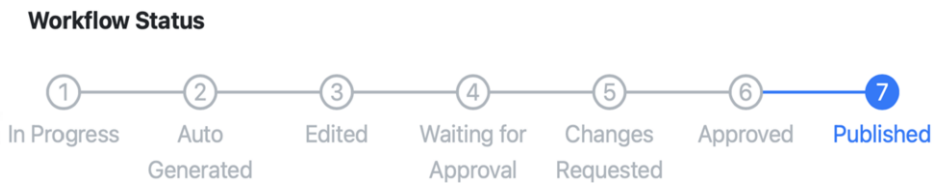


Figure 1. Visualization of workflow status

This article aims to offer an insight into the TransPipe [3] development from design, test and deployment, followed by a brief summary of the tool’s performance since its deployment. The article also discusses the limitations and potentials for upcoming iterations.

2. Methods

TransPipe was developed to manage the translation and transcription process that was, previously, a combination of online tools and manual work without medical or WHO-specific terminology banks. OpenWHO conducted six iterations to test the transcription and translation accuracy with the new external solution provider, and to fine-tune the

workflow with TransPipe. Feedback generated from testing was used regularly for improvements throughout the project. As mentioned before, TransPipe itself is not the actual service provider, but the connection between the learning platform and several external services. TransPipe integrates with OpenWHO through a custom Application Programming Interface (API) and thereby has access to all course main videos of the learning platform. Upon request, TransPipe uses the following two main services offered by external providers to process the videos and create the subtitles:

1. automated speech-to-text transformation to generate video subtitles in the original language,
2. automated translation from one language to another to additionally provide video subtitles in additional languages.

During the implementation of TransPipe, WHO and HPI collaborated with service providers to improve the quality of their services. At a certain point, WHO focused on one of these service providers and tested the accuracy of transcribing English audio from OpenWHO course videos and the subsequent transcript translation into the UN's official languages (Arabic, Chinese, French, Russian, and Spanish) and Portuguese. The tests highlighted differences in accuracy for different languages, with Latin script languages (French, Spanish, and Portuguese) being translated with 70-90% accuracy. At the same time, testers found Arabic, Chinese, and Russian translations to be around 60% accurate. However, subsequent tests found improved accuracy across the board as the default terminology and vocabulary bank expanded with the tool's repeated use. At the upper end of the spectrum, the translation accuracy rivaled competitor transcription tools. Accuracy could continue to be improved with the development of a custom terminology bank populated with WHO-specific terminology and vocabulary, which was impossible with existing competitors at the time.

Once WHO had ascertained a satisfactory accuracy level through internal testing, OpenWHO's platform service provider, the Hasso Plattner Institute (HPI), included the selected external transcription and translation solution in TransPipe's list of supported services. The selected external solution also includes a terminology tab to enable users to identify the errors caused by the machine and save the correct terminology in the repository for future use. Connecting TransPipe to the OpenWHO platform via a dedicated API finalized the deep integration of the new process within OpenWHO's workflow. The ultimate goal was to provide learning materials in different languages, achieving sufficient accuracy to reduce the turnaround time on each piece of content from a few days to a couple of hours. The newly created TransPipe tool also improves collaboration within a team of language experts on the subtitles. It allows assigning specific team members to focus on particular languages of the videos and can automatically notify others when the workflow may proceed. Today, TransPipe facilitates the transcription and translation of all OpenWHO's transcripts and subtitles.

3. Results

By connecting the existing services and filling in the gaps, TransPipe has fully replaced the different manual steps of the previous workflow with automated or semi-automated processes. Particularly, it has eliminated the need to use error-prone workflows including desktop tools, and provides a uniform layer atop of all transcription and translation services that have been used before. Because of this, it has become a one-stop shop for course administrators and editors (translators and QA managers), who can now all work

with the same web-based application. The languages in which the course is to be translated, and the required transcription and translation services can now be configured. Then TransPipe triggers the transcription by invoking the APIs of the configured transcription service, and translators and QA managers can be appointed for each selected language.

Translation for the other languages can be triggered. Editors are notified upon completion via email, which allows them to log into the TransPipe interface and start the quality control process. TransPipe always provides a visualization of the video and course status as all transcription and translation tools offer asynchronous processes to allow some time to return the results. The same applies for the human QA processes. Once the authorized person in charge decides that the quality is sufficient, one more click is needed to publish the subtitles in all translations for all course videos. Figure 2 compares the subtitle generation process with and without TransPipe.

As TransPipe was rolled out, it enabled a process for publishing fully automated multilingual subtitles on OpenWHO, requiring very little manual human input. By the end of 2022, the TransPipe tool will have contributed to the transcription of nearly 4,700 minutes of video content and the translation of 1,050,700 characters in the context of video subtitles on OpenWHO. As of November 2022, TransPipe has primarily been used to transcribe a total of 4,250 minutes of English audio whereas Arabic has amounted to some 100 minutes.

TransPipe has been used to translate English subtitles into other official UN languages, such as Arabic, French, Spanish and Russian, as well as Italian, Portuguese, and Turkish. In addition, it is possible to use TransPipe to translate into seven other target languages (18 languages in total), including some for which the reliable acquisition of translators has proven difficult, such as Telugu and Tamil. The connected external service offers an even more significant number of source and target languages (18 and 51, respectively), which WHO and HPI could enable on TransPipe in the future. Being powered by machine learning, the translation provision integrated with TransPipe continuously augments its vocabulary bank to improve transcription and translation accuracy.

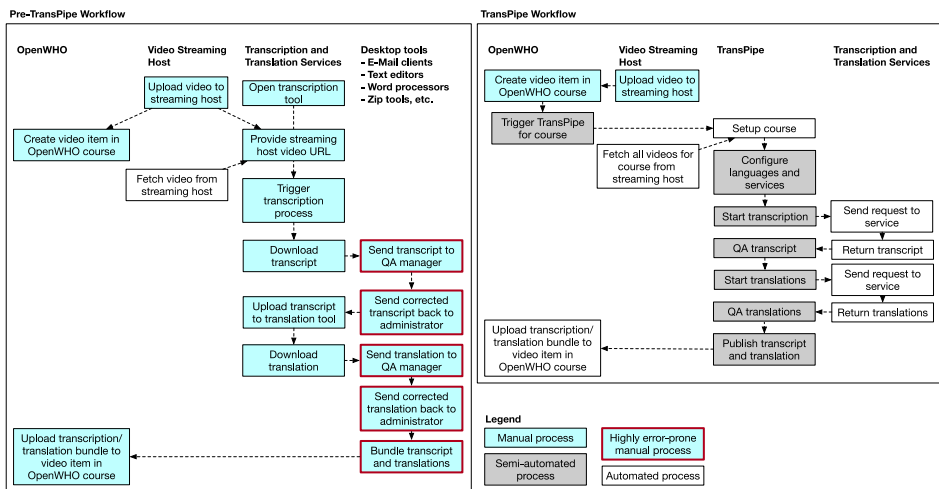


Figure 2. Pre-TransPipe vs. TransPipe process

4. Discussion

Automated transcription and translation have enormous potential as public health learning tools. It has allowed the near-simultaneous availability of video subtitles on OpenWHO in many languages. In addition, the time saved by automatic transcription and translation can hardly be understated compared to doing the same manually. However, there are also limitations. For example, for under-resourced languages, in particular, the validation process could still be challenging despite the improved efficiency [4]. The missing evaluation of the linguistic complexity in the source language also raises questions about quality consistency and learning standards. Continuous research and effort are required to fully realize its potential, including the continued creation of tailored training datasets for optimal transcription and translation accuracy and solidifying a community of practice for testing and rollout of novel tools. Once these tools are well-established and provide satisfying results, a seamless integration into existing contexts and applications is required to leverage the full potential. With the work at hand, we perform the first step in this direction. However, further questions arise, for example regarding the maintenance of user-defined terminology across individual videos or the use of the manually corrected transcripts and translations for supervised machine learning.

5. Conclusion

Technology must serve humanity and keep pace with our needs. The Covid-19 pandemic has provided unprecedented opportunities to innovate and make the fruits of technology available to increasingly large numbers of people. The principle of equity is essential in all work to achieve the broadest of public health goals, especially in health emergencies. For WHO, the OpenWHO platform and technological advances must help remove barriers such as cost, language and digital access so that anyone anywhere can access life-saving knowledge.

References

- [1] Che X, Luo S, Yang H, Meinel C. Automatic Lecture Subtitle Generation and How It Helps. ICALT: IEEE 17th International Conference on Advanced Learning Technologies. 2017; 34-38, doi: 10.1109/ICALT.2017.11.
- [2] Zhao U, Samo G, Utunen H, Gamhewage G, Stucke O. Evaluating Complexity of Digital Learning in a Multilingual Context: A Cross-Linguistic Study on WHO's Emergency Learning Platform. ICIMTH: *Studies in Health Technology and Informatics*. 2021; 281:516-517, doi: 10.3233/SHTI210222 2021.
- [3] Bethge J, Serth S, Staubitz T, Wuttke T, Nordemann O, Das PP, Meinel C. TransPipe: A Pipeline for Automated Transcription and Translation of Videos. EMOOCs: 7th European MOOC Stakeholder Summit. 2021; 79-94, doi: 10.25932/publishup-51030.
- [4] Besacier L, Barnard E, Karpov A, Schultz T. Automatic speech recognition for under-resourced languages: A survey. *Speech Communication*. 2014; 56:85-100, doi: 10.1016/j.specom.2013.07.008.