

2018

ISSN 1433-2620 > B 43362 >> 22. Jahrgang >>> www.digitalproduction.com

Publiziert von DETAIL Business Information GmbH

Deutschland  
Österreich  
Schweiz

€ 17,70  
€ 19,-  
sfr 23,-

1

DIGITAL  
PRODUCTION

# DIGITAL PRODUCTION

MAGAZIN FÜR DIGITALE MEDIENPRODUKTION

JANUAR | FEBRUAR 01:2018



## Finishing

IMF, HDR, Retusche  
und mehr fürs Finale

## Crash! Bam! Pft!

Foleys, Fairlight und Libr-  
aries im Audioschwerpunkt

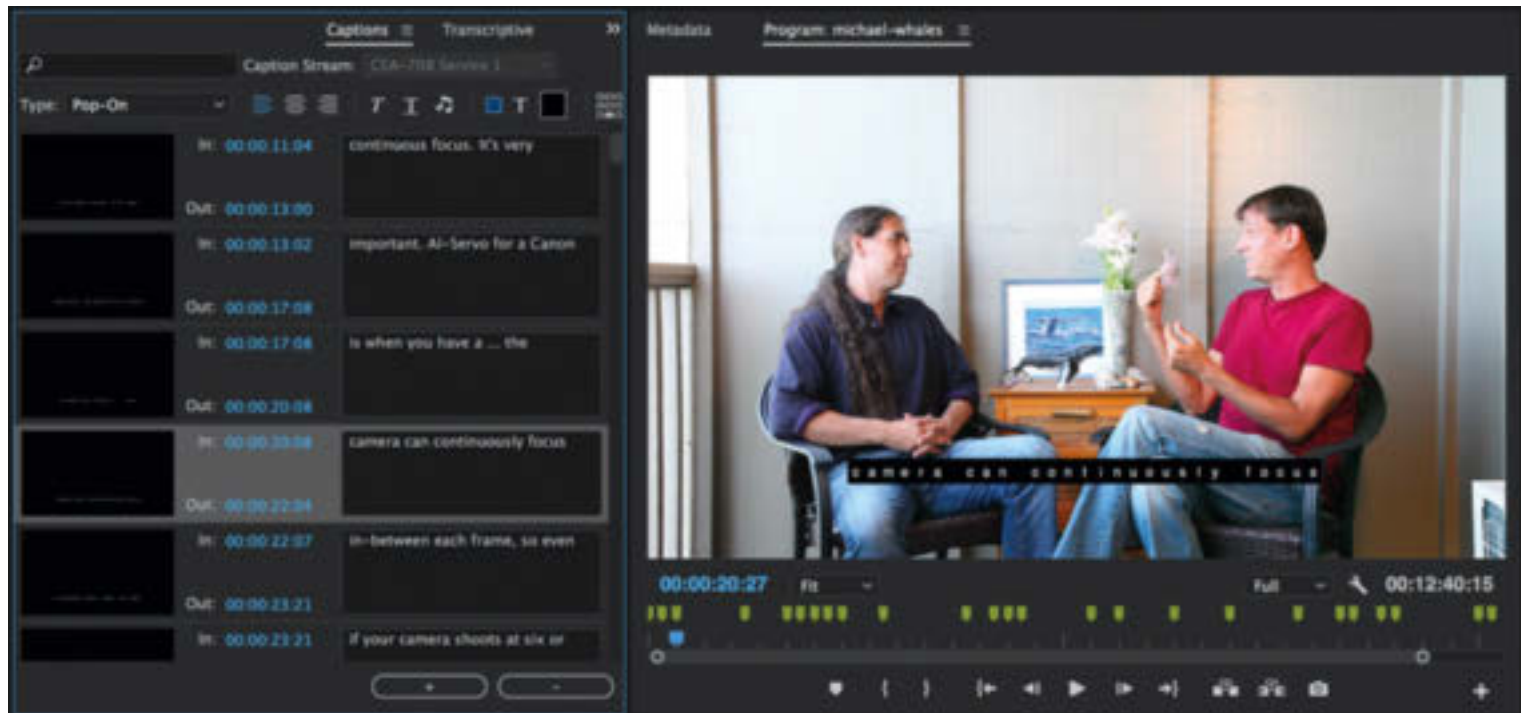
## Workshops!

Clarisse, Fusion, Flame,  
Lumion, Tracking ...



4 194336 217709





The Transcription in Premiere's native interface

## Transcriptive – when AI helps with your editing

One of the most annoying jobs is the transcription of video – and it is often offloaded to interns and apprentices. But what if you could have the so-called artificial intelligence, that everyone is talking about, help you on that? You do not have to wonder anymore, because there is a Premiere Pro plug-in, that promises to do just that.

von Béla Beier

One of the more interesting releases of this year's IBC was "Transcriptive", a plug-in for Adobe Premiere CC, that takes the audio track and loops it through speech recognition. You get the choice of two AIs, namely IBM's Watson and the "Speechmatics" Service – all from within Premiere. If you are wondering if that works for the German version – Andreas Zerr has tested it. Read all about it on page 74.

The plug-in costs 299 US-Dollars, plus you have to pay for the minutes of transcription, depending on how much of it you need, and what quality the audio has. After the plug-in has done its job, you have a transcription including however many speakers actually appear in the video, even though with shifting intonations and pauses it sometimes recognizes a few more people. To hear what is what, we talked to Jim Tierney from Digital Anarchy.

**DP: How did you get the idea to use AI-Speech Recognition for Transcription?**

**Jim Tierney:** Like a lot of our products, it came from wanting to solve one of our own problems: creating subtitles for our Youtube tutorials. At the time, Youtube automatic subtitles were awful and mostly useless (they have improved since then but generally still need to be edited). So I started doing a little research and realized that machine learning was getting better and some of the speech-to-text algorithms were quite good.

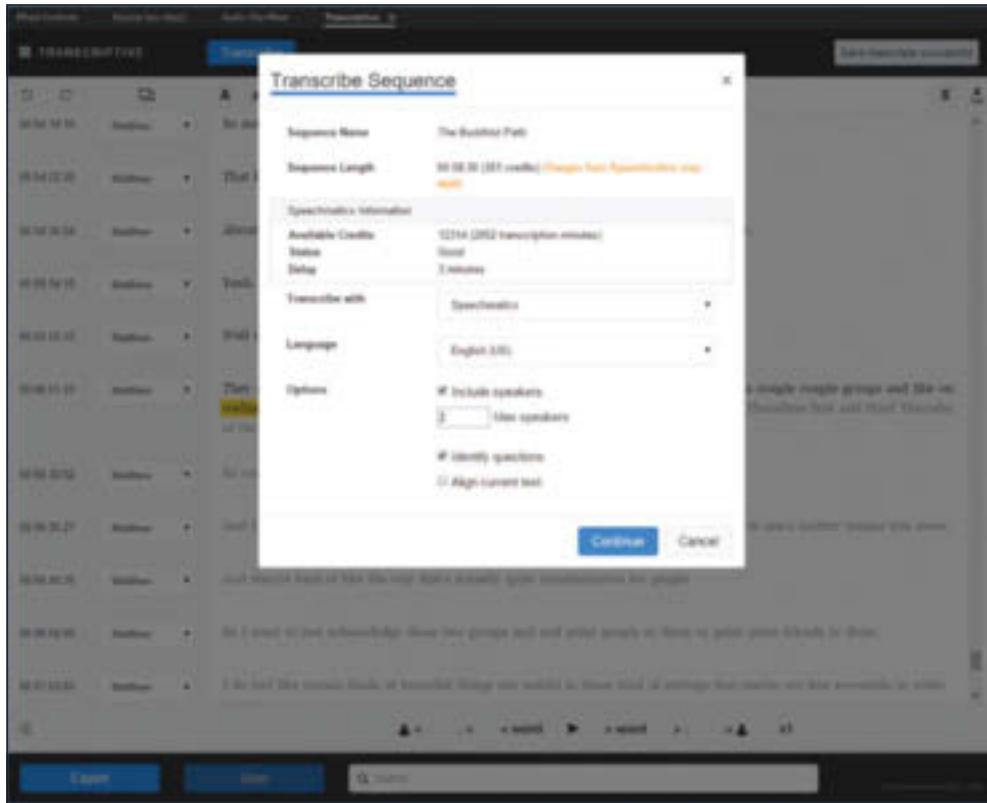
**DP: How close is it to recognizing language, and where are the limitations?**

**Jim Tierney:** It is very good in most cases. Of course, the cleaner the audio, the better spoken the talent is, the less background noise ... all these will make it more accurate. For example, IBM's Watson is good on audio without background noise, but the

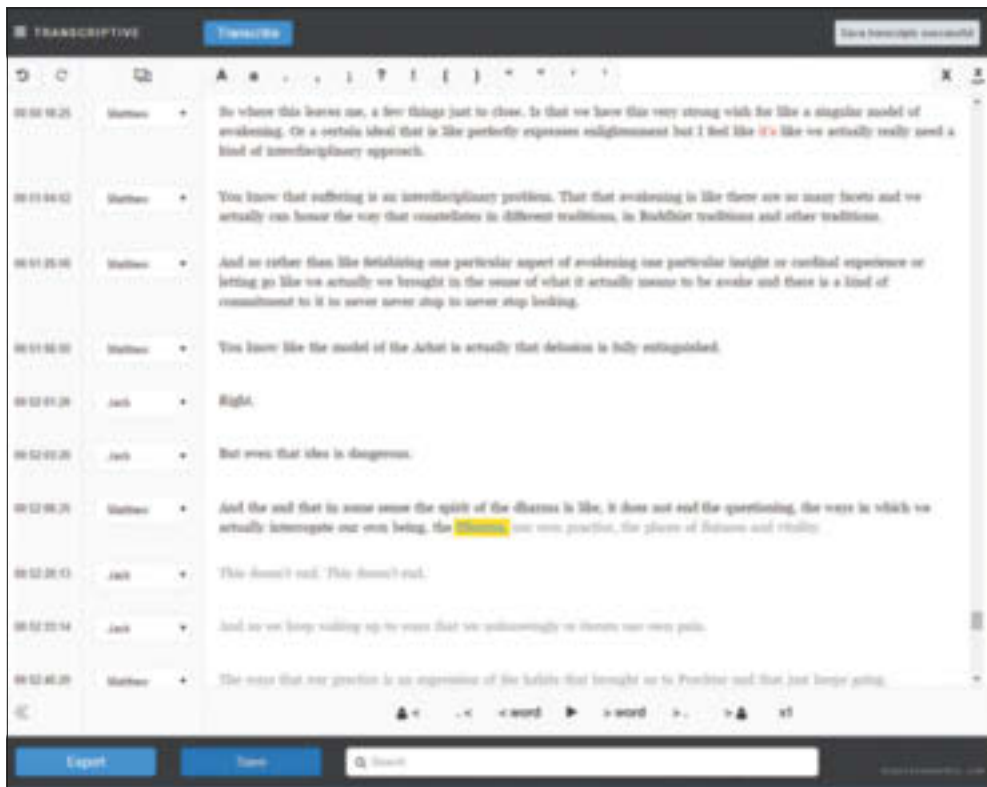
reason we recommend Speechmatics is that it works great with or without background noise (within reason). For the more common languages (English, German, French, etc.) it works great. For less common languages (Swedish, Portuguese, etc.) it works OK and is improving. Speechmatics supports 28 languages, Watson supports 9. Transcriptive lets you try both Watson and Speechmatics and use the one that works best for your needs – just give it a try. By the way: you can download a free trial version on [digital-anarchy.com](http://digital-anarchy.com).

**DP: What is the minimum quality required for something to feed to the AI?**

**Jim Tierney:** It is tough to say. It varies depending on background noise, speaker enunciation, etc. As mentioned, Speechmatics will do better on lower quality recordings than Watson. As long as someone



The setup for the Transcriptive-plug-in. You choose which AI, language and so forth.



The result: speakers are identified (with a little bit of editing), and along with timecode and a bit of text editing you have a decent transcription, without spending days flipping through the frames.

is properly mic'd it works well, even if there's a lot of background noise. The more 'off mic' someone is, the worse it is.

**DP: Since there are two AIs to choose from: which one is better?**

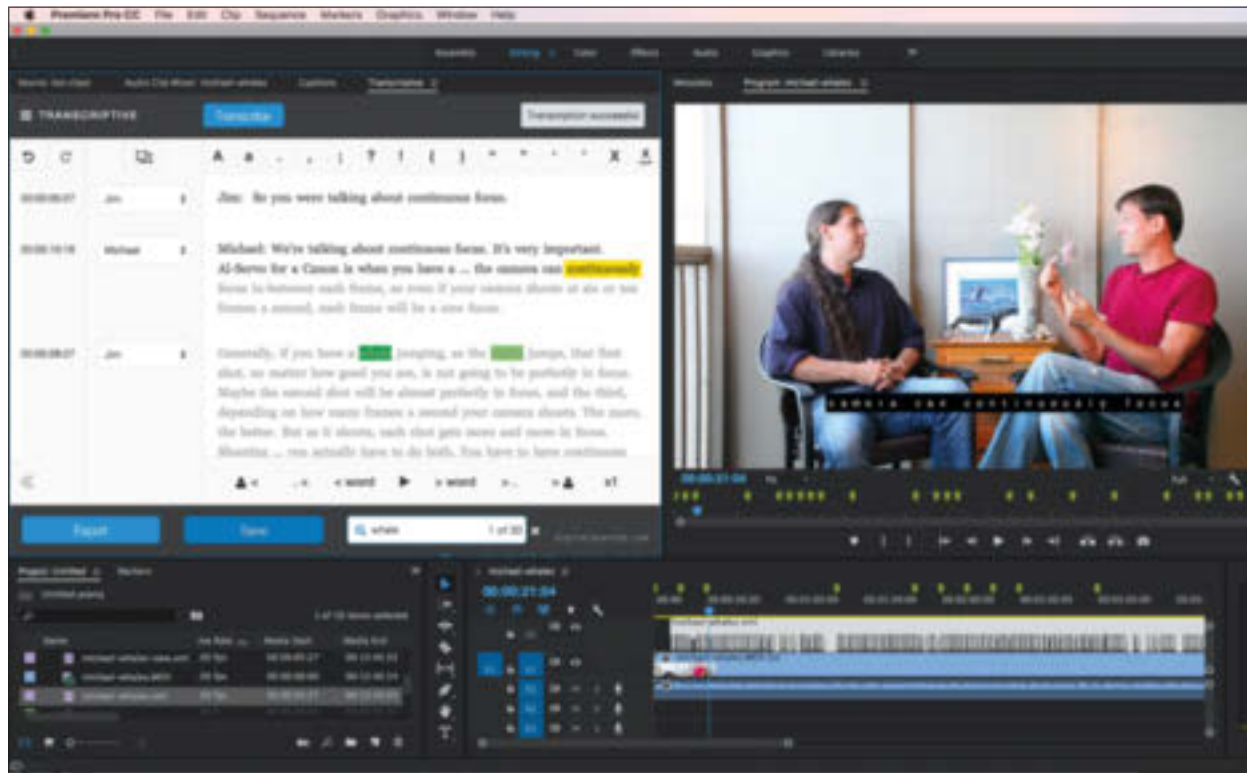
**Jim Tierney:** As mentioned, Speechmatics, hands down. But Watson is good if you have very clean, well recorded audio or you just

need the transcript for video search.

**DP: Will it be available for other editors? For example Resolve, MC or Sony Vegas?**

**Jim Tierney:** The reason why we did it for Premiere Pro first was that the API allows us to create a fully integrated solution. Transcriptive is a web app that lives in Premiere which means we can also control or be

affected by Premiere's functions. For example, if you click on a word in the Transcriptive panel, the playhead in your sequence's timeline will jump to where that word was spoken. Other host apps don't allow us to do that, so we can't do as integrated of a solution. We do plan on doing a standalone app that will work with FCP, Resolve, etc. by sharing metadata. Which is cool, but it is not



Transcriptive within Premiere – you can search for words and phrases, and jump to the exact point in the video.

as nice of a solution as what we have done with Premiere. But that is coming around the beginning of next year.

**DP: And how many corrections are necessary?**

**Jim Tierney:** It is about 90% accurate, give or take. So probably about 1 correction per 10 words. But sometimes those corrections are just adding a period or comma. Transcriptive is essentially a word processor for transcripts. So there are a lot of ways that we use keyboard shortcuts and other tools to speed up the editing process. It is much faster than if you were doing it in a normal text editing application.

**DP: Can the script be exported to other applications further down the pipeline?**

**Jim Tierney:** You can do it now. You can export .srt, .vtt, SMPTE Timed Text, .scc, .mcc and Plain-Text-formats, all of which can be imported into tools like MacCaption, other video editing apps or used for Youtube subtitles.

**DP: Do you have any influence over the pricing for Watson and Speechmatics?**

**Jim Tierney:** Unfortunately no. But we also don't add any costs to it. So the user pays whatever the services charge. We don't make any money on the minutes used, which is why we included Watson. Yes, it is not as accurate, but we figured some users would really benefit from being able to transcribe over 16 hours a month for free. And if another service comes out that we

like, we can add that down the line. We are totally agnostic about what service people can use as we don't make money on it. We just sell the plug-in. It is a little more expensive at the outset, but long term, the cost is much lower. For example, if you transcribe 50 hours of video with a service charging 37 cents per minute, that's 1100 Dollars more than what you would pay with Transcriptive and Speechmatics (7 cents per minute).

**DP: Are there flat rates or volume tariffs available?**

**Jim Tierney:** Not for Watson. You get 1000 minutes free, then it is 2 cents per minute. Speechmatics lets you buy in bulk and if you want a lot of minutes, you can negotiate with them and really get the cost down.

**DP: Transcription is often offloaded to an assistant: Is there a standalone version or an assist station without working from within Premiere available?**

**Jim Tierney:** Not at this time, but that is coming soon, along with some other very cool functionalities.

**DP: What has been the feedback so far, and which new features are you working on?**

**Jim Tierney:** The feedback has been great. A lot of folks have been really blown away by it. We are about to release 1.0.1 which improves the search engine and adds more export formats (and fixes some bugs because ... well, 1.0 software). We are adding Find and Replace soon. That's been something a lot of folks are asking for and something

any word processor should have. We've got some very cool things in development that I think people are going to be excited about, like editing the transcript and automatically generating a rough cut based on the paper edits. But those are a couple of months away, so I'll wait until they are in beta (and I'm 100% sure they will work) to start talking them up too much.

**DP: What are the next steps for speech analysis for the coming years?**

**Jim Tierney:** I think it's just a matter of improving accuracy and adding more languages. There is definitely room for improvement with some of the lesser used languages. I'd also like to see the ability to customize the machine learning algorithms. Right now there's no way of adding words. So if you have an industry with a lot of jargon, the accuracy is lower. Having a way of training certain words would be great. That is coming, but I'm not sure how fast. Overall I think AI/Machine Learning is really interesting and there are a lot of cool things that can be done with it. We're definitely looking into ways we can bring other tools using it to help video editors with their work. Stay tuned! > ei



Jim Tierney is the President and Co-Founder of Digital Anarchy and has been working on visual effects software since he was at MetaTools in 1995. He started Digital Anarchy in 2001 because he loves creating cool tools for video editors and VFX and because he hates cubicles and other corporate nonsense. For insight into life at Digital Anarchy:  
<https://youtu.be/-2eBuNU5xTI>