

Inside: *TVBEurope's* Sports Broadcast conference preview – page 6

TVBEUROPE



United Business Media

Europe's television technology business magazine

www.tvbeurope.com

NOVEMBER 2008 £5.00/€8.00/\$10.00

Companies like SysMedia are aiming at the 'Holy Grail' of subtitling

Breaking the sound barrier

Guest Opinion

By Andrew Lambourne
Chief Executive Officer,
SysMedia

Gone are the days when certain programmes were designated 'too hard to subtitle' — the coverage is now required on everything: live and recorded, rehearsed and real-time. So how can you achieve this?

The classic approach to subtitling a recorded programme is to use a conventional keyboard to retype the dialogue, timing it as you go. Modern methods

borrowed from live subtitling now mean that realtime transcription techniques (such as re-speaking to trained speech recognition engines) are now used to capture the text — and indeed to capture the timing as well — all in one pass. If a script is available, text-to-speech alignment tools can be used to time the script to the audio in order to generate first-pass timecoded subtitles for manual review.

The drivers behind this convergence are simple: to increase productivity and reduce costs, and to make a given team more flexible by being able to address a wider range of

programme genres. Capitalising on economies of scale like this in driving down costs is one reason why subtitling operations can become more successful as they grow.

Convergence is also happening with the remote-working approach. Translation subtitling for DVD has been the stronghold of the freelance subtitler: given a timed master file, the freelancer translates the subtitle text and delivers one of the 30 or so language files that are used to generate the subtitle DVD assets.

This can be done at home — ideally with a browse-resolution

copy of the media file securely delivered over an IP link. This approach is now being adopted more widely for live subtitle production. A remote subtitler creates and submits realtime subtitle text during a live broadcast using a Stenographic keyboard or a speech recognition system.

erate subtitle timing or transmission cues)

- Automated transcription (ie the production of a transcript in realtime from the audio of a presenter or interviewee, or the production of a transcript in non-real time from a media file)

- Automated editing (ie the analysis of the grammar and syntax of a sentence in order to propose a reduction in word-count so as give subtitles adequate display time)

- Automated translation (ie the conversion of a subtitle script in a "master language" into an equivalent script in a target language)

Subtitling remains a cost drain for broadcasters, and the more required the more important it becomes to maximise productivity while still achieving quality

For some broadcasters, the quality of realtime generated subtitles has now become sufficiently good for a decision to be made to use this method during the whole of a live programme.

Speech and language tools, coupled with advanced audio processing software, underpin many of the recent advances in subtitling technology and productivity. Companies such as SysMedia are aiming at the 'Holy Grail' of subtitling: maximum automation/acceptable quality. So far this goal has been achieved in part with more to come.

There are four main areas of automation research:

- Automated timing (time-aligning a manually created script or transcript with recorded or realtime audio so as to gen-

Automated timing

Timing automation uses speech alignment technology to match the phonetics of an audio track with the phonetic structure of a subtitle text to determine the times at which the corresponding words were spoken, therefore to assign subtitle times. A simple challenge? In reality, pretty complex.

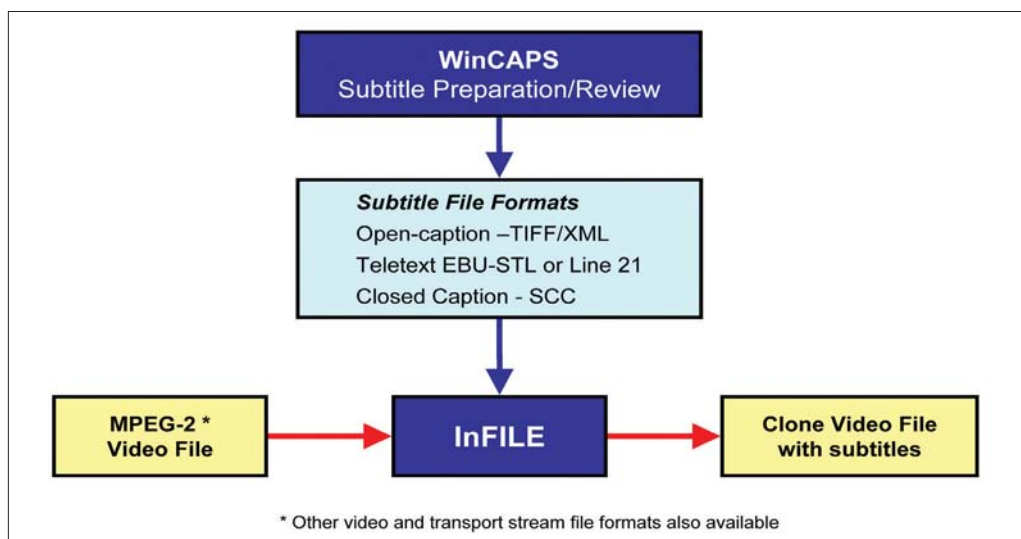
Fully mixed audio tracks generally contain speech surrounded by other noise — music, effects and background. Producers and broadcasters wishing to benefit from alignment technology should seriously consider making unmixed dialogue tracks available for subtitle timing; the benefit is significant.

The problem can be lessened using other tools to attempt to 'gate out' the music and effects, but such tools can't always cope. Nevertheless, auto-alignment tools are available for use with suitable programmes and deliver valuable productivity benefits. Given a spoken documentary-style programme and a script, software can align script-to-dialogue and deliver subtitle timings in significantly less than programme run-time.

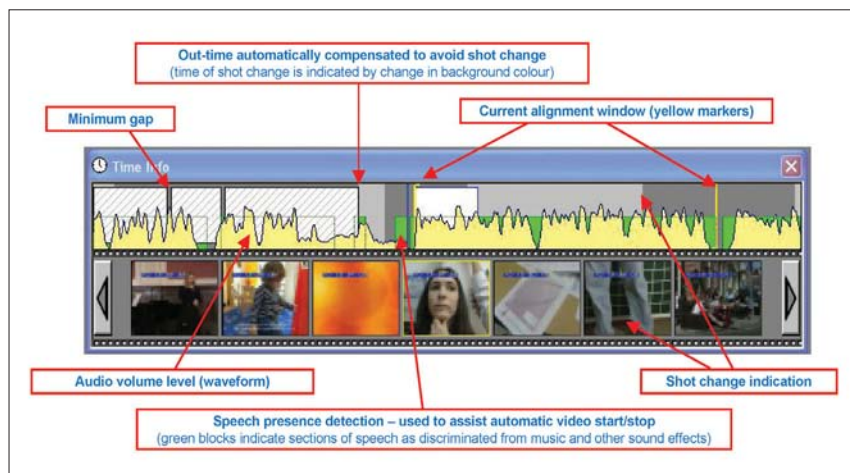
The area where automated timing has really yielded benefit is in the synchronisation of a subtitle script to the voice of a TV presenter in realtime during news broadcasts. In this situation, we can make the software follow what the presenter is saying and transmit the corresponding subtitle script in synchrony and fully automatically, with no previous training.

Automated transcription

Automated conversion of unconstrained speech-to-text in realtime, and at 100% accuracy for any speaker without pre-training of the system, remains a distant ideal. Great progress has been made in situations where some of the constraints can be relaxed. If the restriction on no advance training can be removed, and the system is given the chance to build a speech model representing the characteristics of the voice and pronunciation



Example workflow for SysMedia's InFILE software utility designed to insert subtitles into a broadcast quality video file or transport stream



AutoTime is a major productivity boost for subtitle authoring, saving considerable time for broadcasters by timing text to the soundtrack

patterns of a given speaker, significant benefit is achieved and rates of recognition accuracy may rise by 10% from 60-70% towards 80% or more.

Nevertheless, further restrictions must currently be imposed in order to boost accuracy to more useful levels. If speech is clear, well-formed and uses a vocabulary which is known in advance, results are better than for entirely unconstrained speech. Nevertheless, there still remains the problem of inserting punctuation; a challenging task.

Current generation mass-market speech recognition tools do not have an explicit awareness of grammar. Moving to a grammar-based recognition strategy, and then adding contextual awareness to improve recognition accuracy, requires engines that are still in the research labs.

Software-assisted subtitling must achieve levels of accuracy that are useful for broadcast purposes: 96-97% or more for live subtitles, 90% or more for offline transcripts that will be tidied manually. The current method of choice (where suitable engines exist) is to use a re-speaker who speaks in a constrained and precise way and who has carefully trained a speech recogniser to their voice.

Although 97% accuracy or more can be achieved, this does depend on pre-preparation of vocabulary: if the speech engine is not aware of the names of people and places that may feature in a piece, accuracy falls down badly. It also depends on the language: French for example is a challenge, and special techniques have to be used accurately to transcribe unvoiced word endings which otherwise significantly affect quality.

Customers across Europe are already using re-speaking techniques to generate live subtitles, and are also beginning to use them to generate timed transcripts for the purposes of offline subtitling. Moving to fully automated transcription is slower since results depend heavily on audio dialogue clarity and vocabulary range. Developments in this area will be worth watching, since the automated transcript comes with the benefit of timing information as well. Meanwhile, the re-speaking approach can save considerable time.

Automated editing and translation

Here the challenge very much moves from the technical to the linguistic. Work has been done in both these areas to seek to demonstrate whether it is possible to automate the compression of text into available display time, and/or to convert a source text into a target language text without human intervention.

Both challenges are constrained by the normal conventions of subtitling that ensure

that the subtitles are useful to the intended audience. In the case of editing, the normal convention is to 'red pencil' edit where possible, so that the basic flow and structure of the sentence, and the key words in it, are not altered. This is so hard-of-hearing viewers who may have some access to the audio, or language learners, are not misled or confused by subtitles that have effectively been rewritten and use entirely different words or grammatical structure. It is also normally quicker to sub-edit than to rewrite.

In the case of translation, the aim is to preserve the linguistic quality of the original, and to provide a translation that is of comparable richness in the target language. As with transcription, performing this process manually relies on a deep semantic appreciation of the material being processed, which cannot yet be achieved by mechanistic tools.

There will be a tipping point at which the time taken to fix up a machine-edit or machine-translation will become less than the time taken to perform the job manually. In the meantime, such tools can offer useful input to the subtitler, to use or ignore as they judge best, and are more likely to gain acceptance when presented as emerging expert systems that can learn and adapt — ideally through a means by which 'ideal' manual work is fed back for computer learning.

Closing the loops

From the broadcaster's perspective probably most importantly, which tools convey cost benefits for what genres and types of programme? If the marketplace decides, then broadcasters and subtitling agencies must start to experiment with these new tools and realise their potential benefits.

From a regulatory perspective, if viewers cannot read or make sense of the subtitles as a replacement to the audio track, or the quality of grammar or spelling or content is poor, or the timing is such that the subtitles do not blend with the total experience, then automation has unacceptably compromised quality.

The tipping point is the balance between commercial pressures and the potential damage to the brand: it could be argued that even though subtitles are used only by a minority, the quality measures applied to them ought to be comparable with those applied to other aspects of the production. For this to be meaningful, a distinction needs to be drawn between technical quality and editorial quality.

Subtitling remains a cost drain for broadcasters, and the more that is required the more important it becomes to maximise productivity while still achieving quality.