

Re-speaking the truth

ANDREW LAMBOURNE, CEO, SysMedia, reports.

Broadcasters have generally viewed providing optional subtitles for the hearing impaired as being costly and complicated, for little return. However in the US, the UK and now much of the rest of Europe the stick of legislation is now being used to back up the carrot of increased audience figures: accessibility is firmly on the agenda, driving the requirement to provide subtitles towards 100% of programming output.

Clearly this presents a challenge as well as an opportunity. Deaf and hard-of-hearing viewers want to watch the same mix of programmes as anybody else, so optional subtitles need to cover the spectrum of programming. It is easy and inexpensive (a

few hundred Euros per programme hour) to prepare subtitles for a pre-recorded programme that is 'in the can' a few days in advance of transmission. But doing the same job for a live programme? Many broadcasters balk at this challenge, fearing cost and complexity as well as suspecting that quality will be poor.

The good news is that technology has moved on, and whereas it used to be necessary to employ specialist (and scarce) real-time keyboard operators using complicated devices such as Stenograph or Velotype machines to generate text in real time at or near the speed of speech, in a growing number of territories this is no longer necessary. The secret is to use speech

recognition technology plus far more readily available and easily trained staff to re-speak or 'parrot' the programme soundtrack into a computer that generates an accurate transcript and turns it into subtitles.

Numerous services are now on-air seven days a week producing subtitles for live programming using speech recognition technology. Trained operators generate services that would otherwise have been prohibitively expensive. The benefits are real with the hearing impaired population variously quoted as being between 6-8% of any audience, it is natural to expect that as television services are made more accessible to the deaf, the viewership will rise.

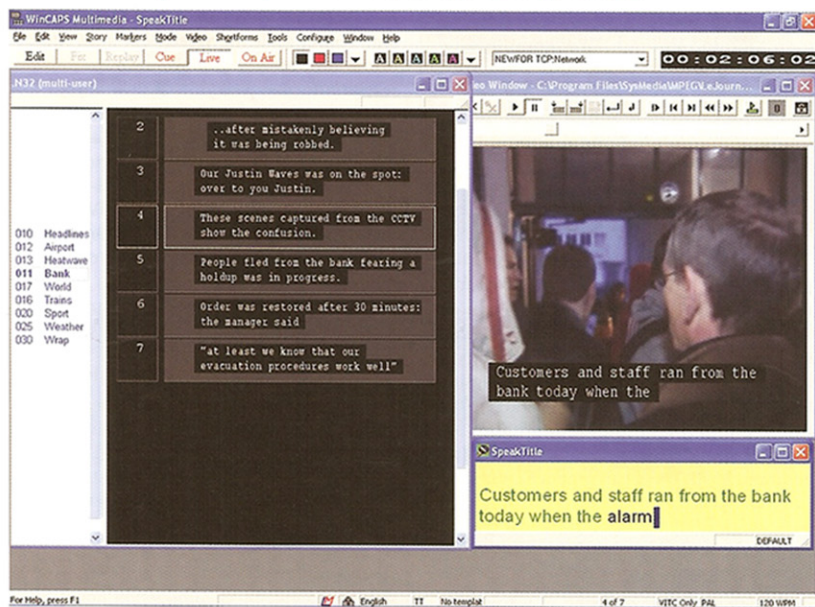
So with the legislative pressure meaning that broadcasters can no longer ignore the need to increase subtitling output across all content types, a fresh look needs to be taken at the available options. Despite perceptions to the contrary, subtitling live programming can be cheaper hour-per-hour than subtitling pre-recorded material, and re-speaking is an important part of the reason why.

Live subtitling options

Subtitling for the hearing impaired in Europe started in the late 1970s, and as regular services emerged, they were restricted to the easier pre-recorded material. Occasional live one-offs were undertaken on an experimental basis using a variety of specialised real-time text generation systems: dual keyboards, Stenograph machines, Velotypes. As much as anything, these early services were grappling with the editorial issues as well as the technical and production challenges: by definition live subtitling cannot be perfect because there is not time to spend carefully composing, perhaps sub-editing, styling, positioning and timing the texts – and unless a broadcast delay is inserted into the sound/vision path the subtitles are also unavoidably late.

It therefore needs to be understood that creating and delivering subtitles in real time for a truly live unscripted TV programme inevitably and unavoidably involves a degree of compromise. Both the broadcaster and the audience need to understand and accept this, and dialogue is helpful as the techniques are explained and the implications understood. The key question is this: given that a live service cannot be perfect, how good does it need to be to be more acceptable than having nothing at all? Once this is agreed, broadcasters can then move forward

A screenshot of SysMedia's SpeakTITLE re-speaking technology at work.



pragmatically to deliver services where compromises are kept to a sensible minimum, quality is preserved at an acceptable level, and cost is reasonable.

Preparing live subtitles successfully is essentially dependent on two things, each inter-dependent:

- Real-time subtitlers with the editorial skill to decide, in real time, what needs to be included in the subtitles.
- A means by which those subtitlers can generate a real-time transcript of what is required, and deliver it with the minimum of delay.

Having one without the other is useless and one affects the other. For example, the decision about how much text to include depends on the throughput speed that the subtitler is able to achieve with the chosen transcription method. And the choice of transcription method depends on the availability of those who can use it. Setting an editorial style cannot be done without taking account of the available means of achieving the service.

Although Stenography led the way as the engine of choice from the late 1970s in the US, and the 1990s in the UK, it has the significant drawback that operators are scarce in Europe and take two or more years to train. Technical advances in speech recognition systems have solved these problems in territories where high quality recognisers exist: there is a much wider pool of potential operators to draw on, and training time is closer to three months. Hence for more and more broadcasters and service providers, re-speaking is now being seen as the way forwards.

What's involved in re-speaking?

Before anything else is possible, a suitable speech recogniser must be available in the target language. By suitable we mean at least:

- Potentially highly accurate recognition (96% or better).
- Real-time operation with minimal throughput delay.
- Ability to support speech 'macros' that expand to specific words/commands.

It may perhaps come as a surprise that even if such a recogniser is not available off the shelf, such things can be produced where there is a political will to do so, at a cost which (at national level) is insignificant – and with benefits in terms of spin-off application to business and home users generally. SysMedia provided advice during such a project in

Denmark in which Philips and PDC, a Danish development company, produced a brand new Danish vocabulary model based on Speech Magic.

Given a suitable recognition engine, there is always an enrolment process in which the engine learns to adapt its recognition to the particular voice and enunciation of a given operator. This iterative process creates and refines a voice model for that user, which overlays the language model for the language.

Part of the development process also involves building and consolidating one or more dictionaries containing any specialist terms (often people or place names) that will be expected to crop up in the broadcasts, as well as developing methods to deal with homophones where words sound the same but are spelled differently.

Thereafter, on a day-to-day basis, the operator will need to prepare any new vocabulary before a broadcast, undertake a warm-up session, generate on-air subtitles, and review and accuracy issues after the session.

From the technology point of view, a well engineered live subtitling system copes with converting the recognised text into subtitles, decoding speech commands to control colour and style, applying house style rules to tidy the presentation and correct recognition errors, and timing the delivery of the subtitles. Often users choose to present the output in a two-line scrolling format where each new word is added to the right-hand end of the bottom row. This minimises the delay between a word being spoken and seeing it onscreen.

Re-speakers are selected and trained on their ability to get good results from speech recognition systems. They need to understand how the system is working and what it needs: essentially clearly enunciated text that is spoken consistently and at an even pace. On top of this they need the ability to make accurate editorial decisions regarding what to leave in and what can be taken out if the text has to be slightly edited down to reduce speed. The result must be factually accurate, comprehensible, and cover all the key points.

Speed, accuracy and error-handling

Speech rates on television can vary widely, but for live news, current affairs and chat-shows rates of between 180 and 220 words per minute can be experienced. It is also possible for more than one person to be speaking at a time. Putting all this text on screen as verbatim

subtitles generates a very high reading load that may mean the viewer cannot keep up (particularly since deaf people may have below-average reading skills), and may obscure too much of the picture. It may also be impossible in a sensible way during simultaneous speech.

Our advice is normally to reduce the text sufficiently to make it more easily palatable without disenfranchising the viewer by making them feel that they are being patronised by over-editing. Speed reduction to say 160wpm increases the potential accuracy.

A well-trained operator using a good recogniser and mature voice model, with high quality audio equipment, should be able to achieve accuracy of around 97% or better in English. It is currently slightly lower in languages such as French or German that are more demanding (of a speech recogniser) in terms of grammar and agreement of endings. Whilst spelling errors would not be seen (except in the case where similar names have been interchanged), typical speech recognition errors involve leaving out or adding in small words, or using the wrong word or phrase.

It is tempting – particularly given the high standards of accuracy that are demanded and achieved for language translation subtitling in Europe – to demand 100% accuracy in live subtitling. The only way to do this is to make more time available – either by delaying the audio/video, or by delaying the subtitles so that a second operator (or more) can attempt to correct the text. However, taking this step significantly increases cost, can produce unacceptably high delays that can render the subtitles almost useless (even if perfectly spelled), and still cannot guarantee 100% accuracy in near real time.

Other approaches to reduce errors involve tightly integrating the subtitling system with the newsroom scripting system so that in cases where the 'live' content is actually being read from a teleprompt, the subtitlers have it available and can cue it out word-perfect. Similarly, access to upcoming VT packages means that the 'live team' can pre-transcribe these if time permits (using speech recognition technology and then correcting where needed), further diluting the 'truly live' content.

By designing the service to take account of the possibilities opened up by new technology, by making pragmatic decisions about editorial style, and by educating and informing the audience, broadcasters can do a lot to pave the way for cost-effective new live subtitling services based on speech recognition, without opprobrium.