

# Captioning Beyond Compliance: Adding Meaning to Media

Kevin Eler, Ph.D., Automatic Sync Technologies, USA  
[kevin@automaticsync.com](mailto:kevin@automaticsync.com)

Brent Robertson, Automatic Sync Technologies, USA  
[brent@automaticsync.com](mailto:brent@automaticsync.com)

Pat Brogan, Automatic Sync Technologies, USA  
[pat@automaticsync.com](mailto:pat@automaticsync.com)

**Abstract:** Video is becoming more and more popular for use in academic instruction. As universities integrate video into instruction, they struggle to make the media accessible. Disability laws require that most universities make their media accessible to students with disabilities. For video, this means that captions must be added. For podcasts and audio files, a transcript meets the minimal requirements for accessibility. Developing policies, practices and finding budget funds to captioning media in tough economic times stresses many universities. Captioning media is critical to support the deaf and hard of hearing communities. But the benefits of captioning extend far beyond supporting just these viewers. Most viewers of captions are not deaf or hard of hearing. This paper covers benefits of captioning and best practices for captioning.

## Introduction

Tools to capture, edit, publish, and distribute video are advancing rapidly, and as they do video is becoming increasingly popular in academic instruction. A recent study on the use of video in higher education shows that the video explosion is hitting all aspects of education (New York University, 2009). Cisco projections show that video traffic is contributing to 30% of data transfers in 2009, and by 2013 will be over 60% (Cisco, 2009). Students are expecting to have access to instruction through video, and are creating, uploading and sharing videos as a part of their daily lives.

As universities integrate video into instruction, they struggle to make the media accessible. In the US, a variety of federal accessibility laws mandate most colleges and universities to make their content accessible to students with disabilities. In addition, several states have enacted similar legislation. For video, this means that captions must be added. For podcasts and audio files, a transcript meets the minimal requirements for accessibility. But captioning and transcription is expensive, and the exponential growth in media requiring captioning creates significant competition for the resources that are available. Developing policies, best practices and finding budget funds to caption media needs to be planned early on as colleges and universities develop their media strategies.

Captioning all of the video content produced on a campus is clearly a desirable goal: it not only addresses compliance with accessibility laws, but it offers substantial benefits for all users of the institution's content. However, there are significant barriers to large-scale captioning of video content. These include:

- ▶ *Distributed production:* media production is no longer centralized and controlled by a small group. YouTube, iTunes, and web media mean that any faculty, staff member, or student can publish media.
- ▶ *Plethora of media formats:* there are many different media formats and new ones arrive each year; each format handles captioning in its own unique way – requiring significant technical expertise to understand all of the methods and pitfalls of captioning for each media format.
- ▶ *Disparate technical skill levels:* because media creation and distribution has been decentralized, there is a wide variance in the technical expertise of those charged with creating (and captioning) media.

- ▶ *Speed*: media is frequently delivered "just in time" with no allowance made for making the media accessible. Delaying access to new media until captioning is addressed is often not a palatable solution, so how do you address the issue of timeliness in delivering accessible media?
- ▶ *Scalability*: captioning is a time-intensive process. Even when small operations find a workable workflow that addresses their needs, issues of scalability need to be examined if the solution is to be suitable for a broader deployment.
- ▶ *Cost*: while we found great variance in the cost of implementing captioning solutions, even the most efficient solutions are still very expensive.

While there is no question that these barriers are formidable, the benefits of captioning video material are worth the effort. This paper examines some of those benefits and outlines some practices that help mitigate these barriers.

## The Benefits of Captioning

Captioning media is critical to support the deaf and hard of hearing communities. But the benefits of captioning extend far beyond supporting just these viewers. Indeed, most viewers of captions are not deaf or hard of hearing.

However, it is important to note that while a primary motivator for captioning may be compliance with accessibility legislation, there are in fact many more compelling reasons to caption video content. Ultimately, captioning video content leads to a better educational product, with benefits accruing to all of your viewers. Some of the benefits of captioning include:

- ▶ *Accessibility*: captioning can make your content accessible for viewers with hearing, learning, or other disabilities – allowing you to reach a larger viewing audience.
- ▶ *Alternate Learning Styles*: not all learners absorb material in the same way. Research has shown that universal design techniques make educational material more effective for a broader range of learning styles (Rose and Meyer, 2002).
- ▶ *ESL*: in cases where the student or the lecturer is not a native English speaker (ESL), captioning can improve communication and comprehension.
- ▶ *Improved comprehension*: multi-modal communication improves comprehension and retention for all viewers, regardless of ability. A study conducted at the California State University at San Francisco showed that when captions were made available for videos to a randomly assigned group of viewers, their grade point average was one G.P.A. higher for those with captions (Collins, 2008). In addition, these students exhibited a much more engaged level of interest in the content.
- ▶ *Searchability and Navigability*: for online media, the caption data can easily be used to enable viewers to search and navigate video content via text queries. This allows viewers to more rapidly locate information of interest, and extends the shelf life of your content.
- ▶ *Discoverability*: for electronic media, caption data can be used for "search engine optimization" – enabling search engines (either internal or external) to draw interested viewers to your content.

## Best Practices: Accurate Captioning

Most legislation on captioning is silent on the issue of accuracy, but obviously accuracy is a key factor in determining the usefulness of your captioning effort. Particularly in an educational setting, accurate captioning is critical both for students who rely on it as the primary mode of information as well as for those using it for supplementary re-enforcement. Before one can consider alternate approaches to captioning, it is necessary to consider accuracy issues and their impact on the overall effectiveness of captioning.

The problem is that accuracy is very difficult to measure as any accuracy metric presumes the "correct" answer is known. When the source material is an audio recording, there are a number of factors that can make it very difficult to determine what the "correct" transcript is:

- ▶ *Recording quality*: speakers that are not speaking into a microphone, low recording levels, echo or other room effects, and distortion in the recording can all greatly reduce the intelligibility of a recording and make it difficult or impossible to determine what was originally said.
- ▶ *Noise*: Noise and other background sounds can overpower portions of the original speech, again making it difficult to extract the original text.
- ▶ *Homophones*: it is not always possible to tell from context which is the desired homophone; this is especially true with proper nouns (e.g.: Smith vs. Smyth).
- ▶ *Punctuation*: punctuation is, of course, not present in the audio signal – it is interpreted by the listener; as a result, it can be subjective to determine when the listener used incorrect punctuation.

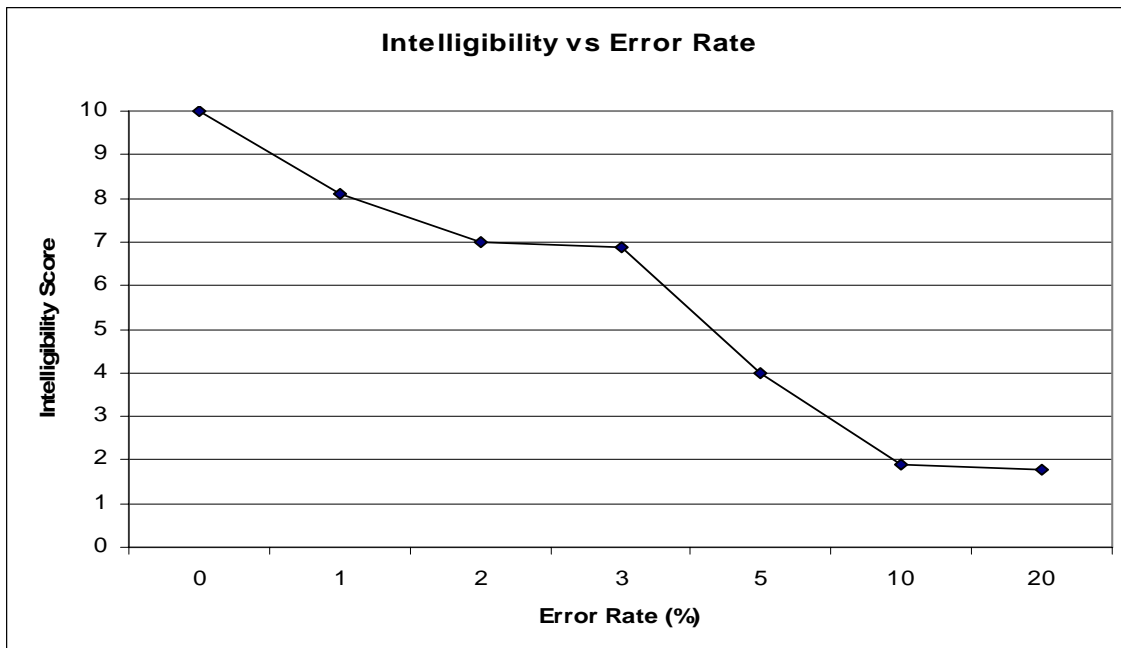
While these issues may not arise in constructed test cases, they are quite prevalent in real-world classroom recordings and can make it quite difficult to determine an accuracy metric.

Nonetheless, if we briefly overlook these issues, the Levenshtein Distance is a typical metric to use to calculate transcript accuracy. This is a metric for measuring the difference between 2 word strings; it is also referred to as the “Edit Distance”. The Levenshtein Distance is the minimum number of operations needed to transform one word string into the other. It is calculated as follows:

$$\text{Word Error Rate} = (S + D + I) / N$$

Where S is the number of word substitutions required, D is the number of word deletions required, I is the number of word insertions required, and N is the total number of words in the transcript.

To get a sense of how transcription accuracy impacts the intelligibility of text, we asked forty test subjects to subjectively rate the intelligibility of a text passage with randomly inserted errors. While it is likely that real-world errors would be different than the random errors that we artificially created, and such different error patterns may have a different impact on intelligibility, this first study was intentionally simple to yield a first estimate of the sensitivity of readers to transcription errors.



Our results show a very rapid decline in intelligibility as the error rate is increased, with most test subjects stating that the text was unusable once the word error rate reached just 5%. By the time the word error rate reached 10%, test subjects reported that they were unable to determine even the topic of the text passage.

Given the size and potential cost of addressing captioning needs across an entire campus, an automated low-cost solution such as speech recognition is certainly attractive and many campuses have attempted to take this path. Unfortunately, while speech recognition technology has improved to the point where some applications of the technology are commercially viable, the general captioning problem remains well beyond the reach of today's technology. Accuracy of speech recognition solutions varies dramatically, depending on the acoustic conditions (mic quality, noise, room acoustics, etc.), content (what the speaker is saying), speaker training (whether the recognizer has been trained to the speaker), and who the speaker is (some speakers deviate too much from the recognizer's internal models making performance especially poor). If the recognizer has not been trained with speech data from the speaker being captioned, typical word error rates are 20% to 40% – well into the "unintelligible zone" from our earlier studies. If the recognizer can first be trained for the speaker being captioned, error rates can fall to as low as 3% to 5% (assuming the audio quality is good – even the trained recognizer will deteriorate rapidly under poor recording conditions). Clearly, the error rates of untrained recognizers were not even remotely close to being a useful captioning solution. Training recognizers to a particular speaker is often not possible or not practical, and even when it can be done, the resulting error rate is still too high to be left uncorrected.

The argument that "something is better than nothing" is often put forward for these sorts of tools, but at best this argument is valid only for "low stakes" content; these tools should not be used for high-stakes content. Speech-to-text tools have low accuracy rates that may be suitable for entertainment videos and for keyword searching. For higher stakes content, such as academic content, accuracy rates typical of speech-to-text tools have not met legal accessibility guidelines, nor have they been acceptable enough to rely on the output to deliver academic education.

## Best Practices: Selecting Material to Caption

With video content growing at an exponential rate and the barriers to captioning significant, most universities are finding it impossible to caption all video content. If anything less than all is to be captioned, the challenge becomes how to select what content warrants captioning. Some obvious first steps are:

- ▶ Determine what media you have; develop a comprehensive inventory of archived media and ongoing production.
- ▶ Make captioning part of the production process so that you are not creating more media that needs to get retrofitted later.
- ▶ Develop policies about what needs to get captioned.

Developing policies to clearly specify what gets captioned and what gets excluded is a challenging task. An approach suggested within the California State University System (Wells, 2010) recommends asking "smart questions" about your media in order to determine eligibility for captioning. These guidelines help determine where resources are spent rather than rigid policy. Some of the "smart questions" include:

- *Is this an Accommodation Request from the Disability Office?* Accommodation requests are the highest priority.
- *Is this media a course requirement?*
- *How often will it be used?* Frequently used media is higher priority than single-use media.
- *Will anyone be tested on the contents of this media?*
- *How many will be using it?*
- *Where is it located?* Publicly accessible content must be captioned.

For most institutions, captioning will be a daunting task. The best practice here boils down to planning: determine the scope of the problem, prioritize the media, then start at the top of the list and work down.

## Best Practices: Funding Captioning

While captioning is an expensive endeavor, the cost is usually not that large when expressed as a percentage of the total cost of media production. The only practical way captioning can become ubiquitous is if it becomes an integral part of the media production process; integral in terms of planning, workflow, and budgeting. This means that

funding responsibility for captioning likely needs to ultimately lie with the producer of the media; the cost of captioning should be treated the same as any other cost associated with the production of the media. In decentralized environments like universities getting to this goal is a process of educating stakeholders, a process that may take some time to permeate the institution.

In the interim, centralized models of funding captioning have been deployed successfully. The California Community College System has implemented the "Distance Education Captioning and Transcription" program which provides centralized funding for the over 100 campuses in the system to caption distance education media (DECT, 2009). While we believe that captioning should be included as an integral part of the production budget, there is no question that programs such as the DECT are important enablers until that happens. The DECT program provides several important services to the participating campuses:

- ▶ Initial vetting and approval of captioning suppliers
- ▶ Centralized administration of the captioning funding
- ▶ A mechanism to apply for projects to be funded
- ▶ Direct payment of caption vendors so that campuses do not have to be involved in the payment process

## Best Practices: Workflow Automation

One key aspect to making captioning pervasive on your campus is to make the process unobtrusive in the production process. If the captioning process causes significant delay in the production or if it creates significant workload, adoption will be slow.

By organizing your media in a centralized place, either on a media server or in a content management system, it is possible to track and control which media pieces are captioned. It is relatively simple to create automated mechanisms that apply captioning rules to select which media needs to be captioned ... and even to take it one step further and submit that media for captioning. Look for captioning solutions that allow electronic submission of content for captioning and that can provide the ability to interface with automated tools. The capabilities will allow your automated content management systems to administer the entire captioning cycle, making the process transparent to the media producers.

As new media production technologies emerge, such as the lecture capture systems now becoming commonplace, look for these systems to not only accommodate captioning, but to provide this same type of automated workflow. Lecture capture systems from Echo360, Mediasite, and Panopto all provide automated captioning workflow as part of their product now.

## Conclusions

Universities and colleges in general, and distance education programs in particular are rapidly expanding their use of video content. As they do, captioning of this content will become both an increasing burden to address and an increasing opportunity for differentiation, so understanding how to tackle this problem is a worthwhile effort. Our work in this field yields a few important points to remember:

- ▶ Compliance legislation is increasing. While there are many other good motivations for captioning your content, this is one that schools need to watch closely.
- ▶ Beyond simple accessibility, there are many good reasons to caption; ultimately captioning can make your education product more effective.
- ▶ Accuracy is key – don't waste resources on solutions that will not provide adequate accuracy. In particular, in our assessment, speech-recognition-based solutions are not viable for captioning at this time.
- ▶ Scalability is also a critical factor in evaluating solution paths. To be viable in the long term, both your captioning solution and the workflow that surrounds it must be scalable.

- ▶ Develop a policy about accessibility in general, and captioning in particular. This is an important step to making captioning an accepted practice on your campus. If you are not able to caption all of your media, develop guidelines on how to prioritize and select media for captioning.
- ▶ Educate media producers about the need to caption their content and the value of doing so. Make captioning part of the production process... and part of the production budget.
- ▶ Make the captioning workflow as unobtrusive as possible; automate as much of the captioning workflow as you can to make it transparent to the media producers.

*About Automatic Sync Technologies:* Automatic Sync Technologies was funded in part by a US Department of Education grant to do research and develop an improved captioning system. Almost 1,500 users today use AST to caption their video, with the majority of users being colleges and universities.

## References

Cisco (2009). *Cisco Visual Networking Index: Forecast and Methodology, 2008–2013*, [http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white\\_paper\\_c11-481360.pdf](http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360.pdf)

Collins, R. (2008). ...And Captions for All?: A Case Study of the Relevance of Using Captions in a College Classroom. *The Garden Newsletter*, May 2008, <http://ctfd.sfsu.edu/sites/default/files/publications/The%20Garden%20-%20Volume%20%20Issue%204%20-%20May%202008.pdf>

DECT (2009). Distance Education Captioning and Transcription program, [http://www.canyons.edu/Offices/Distance\\_Learning/Captioning/](http://www.canyons.edu/Offices/Distance_Learning/Captioning/)

New York University (2009). *Video Use and Higher Education: Options for the Future*. [http://library.nyu.edu/about/Video\\_Use\\_in\\_Higher\\_Education.pdf](http://library.nyu.edu/about/Video_Use_in_Higher_Education.pdf)

Rose, D. & Meyer, A. (2002). *Teaching Every Student in the Digital Age: Universal Design for Learning*. Alexandria, VA: Association for Supervision and Curriculum Development. <http://www.cast.org/teachingeverystudent>

Wells, J. (2010). *Captioning Guidelines: Deciding What to Caption Without a Budget*. Community of Academic Technology Staff (CATS) 2010 Conference, April, 2010. [http://confsys.csuprojects.org/session\\_detail.cfm?sessionID=541&OccurrenceID=458](http://confsys.csuprojects.org/session_detail.cfm?sessionID=541&OccurrenceID=458)