"The Use of Respeaking for the Transcription of Non-Fictional Genres An Exploratory Study", *inTRAlinea* Vol. 19.
Stable URL: http://www.intralinea.org/archive/article/2262

The Use of Respeaking for the Transcription of Non-Fictional Genres
An Exploratory Study

By A. Matamala, P. Romero-Fresco & L. Daniluk (Universitat Autònoma de Barcelona, Universidade de Vigo, Spain & University of Roehampton, UK)

Abstract & Keywords

English:

Transcription is not only a useful tool for audiovisual translation, but also a task that is being increasingly performed by translators in different scenarios. This article presents the results of an experiment in which three transcription methods are compared: manual transcription, respeaking, and revision of a transcript generated by speech recognition. The emphasis is put on respeaking, which is expected to be a useful method to speed up the process of transcription and have a positive impact on the transcribers' experience. Both objective and subjective measures were obtained: on the one hand, the time spent on each task and the output quality based on the NER metrics and on the other, the participants' opinions before and after the task, namely the self-reported effort, boredom, confidence in the accuracy of the transcript and overall quality.

Keywords: audiovisual translation, subtitling, respeaking, speech recognition, transcription

1. Introduction

A written transcript of an audiovisual product may be needed in various scenarios: for instance, audiovisual translators (whether for subtitling, dubbing or voice-over) often require a written script of the audiovisual content they are translating in order to reduce costs and increase translation accuracy and speed. Whereas up until now this job was only performed by professional transcribers, translators are increasingly being asked to produce transcriptions (Cole 2015). As for the notion of accessible filmmaking (Romero-Fresco 2013), which envisages the inclusion of translation and accessibility from production, companies such as Silver River, Endemol or Subtrain in the UK are beginning to employ translators to first transcribe and then translate documentaries. Furthermore, translators are also being hired to provide transcriptions for EU-funded projects on speech recognition-based subtitling such as SAVAS (**http://cordis.europa.eu/fp7/ict/language-technologies/project-savas_en.html**, see below), and transcriptions obtained using different methods (automatic/semi-automatic/manual, professional/non-professional) are also generated to process large speech corpora in research environments.

Professional transcribers often provide high-quality manual transcripts. However, this paper aims to explore whether alternative processes based on speech recognition systems, either automatic or via respeaking, could be successfully used when generating a written transcript of an audiovisual product in a professional environment. The focus will be on non-fictional content that has to be voiced-over in the target language and has not been previously scripted. Our hypothesis is that respeaking could be successfully used to speed up the process of transcription and could have a positive impact on the transcribers' experience.

This exploratory research is linked to the three-year project "ALST-Linguistic and Sensorial Accessibility: Technologies for Voice-over And Audio Description", funded by the Spanish Ministry of Economy and Competitiveness (reference code FFI-2012-31024) in which the implementation of three technologies (speech recognition, machine translation and speech synthesis) in two audiovisual transfer modes (audio description and voice-over) was investigated. The focus in this exploratory study is on speech recognition as applied to the transcription of voice-over, with special emphasis on respeaking.

The article begins with brief definitions of some of the key concepts used in our research, especially of what transcribing means (section 2), and then offers a short overview of previous research on respeaking, mostly as a transcription tool (section 3). It must be stressed that, even though three transcription techniques are tested in the experiment (manual transcription, respeaking, and revision of an automatically-generated transcript), our interest lies mainly in respeaking, since our hypothesis is that this technique will yield better results. The third section is therefore devoted to this technique and does not address transcription via other procedures. Finally, the article describes the experimental set-up of the current test and discusses its results.

2. Transcribing: different needs, different approaches

Transcription is often regarded simply (and simplistically) as the written representation of audible speech. This definition fails to account for the complexity involved in a phenomenon that, as demonstrated by Ochs (1979) in her seminal work, is theoretical in nature, "a selective process reflecting theoretical goals and definitions" (ibid. 44). Indeed, despite the inadequate attention given to transcription in qualitative research (Oliver *et al.* 2005), there is now agreement that transcribing is an interpretive act rather than simply a technical procedure (Bailey 2008). According to Davidson (2009), transcription may thus be defined as follows:

> a representational process (Bucholtz, 2000; Green et al.,1997) that encompasses what is represented in the transcript (e.g., talk, time, nonverbal actions, speaker/hearer relationships, physical orientation, multiple languages, translations); who is representing whom, in what ways, for what purpose, and with what outcome; and how analysts position themselves and their participants in their representations of form, content, and action. (Green et al., 1997, p. 173)

An example of the complexity of transcription is found in projects creating corpora from speech data, as put forward by Bendazzoli and Sandrelli (2007), who state that transcribing was the most challenging task in their study. Some of the aspects that need to be clearly defined when preparing transcription protocols are the level of detail in the annotation (linguistic, paralinguistic, non-verbal and visual information, for instance), the spelling and punctuation conventions to be applied, and the technology to be used (Bailey 2008). Indeed, transcribing implies making choices and this selectivity of the transcripts is always related to the goals of the study, as "it is impossible to record all features of talk and interaction" (Davidson 2009: 38).

Taking into account the aim of our research, transcription is not viewed as a process to generate annotated data for research purposes but as a professional task in which a written text needs to be created from a non-scripted audiovisual source. Hence, participants were expected to follow the standard procedures in the video production industry, which includes a written representation of audible speech, the identification of the person speaking and relevant paralinguistic features such as <laugh> that can facilitate the editing process. No further annotations were requested, as these were the ones a professional audiovisual translator working on a non-fictional content to be voiced-over would need.

Voice-over was indeed the audiovisual transfer mode chosen for the experiment, and can be defined as follows:

> Technique in which a voice offering a translation in a given target language is heard simultaneously on top of the SL voice. As far as the soundtrack of the original program is concerned, the volume is reduced to a low level that can still be heard in the background when the translation is being read. It is common practice to allow the viewer to hear the original speech in the foreign language at the onset of the speech and to reduce subsequently the volume of the original so that the translated speech can be inserted. The translation usually finishes several seconds before the foreign language speech does, the sound of the original is raised again to a normal volume and the viewer can hear once more the original speech. (Díaz-Cintas and Orero 2006:473)

Voice-over is used in many countries to revoice non-fictional content, but it is also used to revoice fictional genres in Eastern European countries such as Poland. Although many voice-over translators work without a script, the availability of an accurate transcript reduces the time and increases the accuracy of the translations (Franco *et al.* 2010).

As said before, transcription of non-fictional content for voice-over has been generally performed manually but two alternative systems are investigated in this research: respeaking and revision of automatic speech recognition transcript. A brief definition of both is offered next.

Respeaking is defined by Romero-Fresco as a technique in which a respeaker listens to the original sound of a live programme or event and respeaks it, including punctuation marks and some specific features for the deaf and hard-of-hearing audience, to speech recognition software, which turns the recognized utterances into subtitles displayed on the screen with the shortest possible delay. Respeaking has become one of the preferred methods for creating live subtitles, although other methods have been used in the past and are sometimes still used: standard QWERTY keyboards, Velotype, tandem or rotation keyboards, and stenotype (Romero-Fresco 2011:1-13). However, in this article we consider respeaking a technique in which the original verbal aural elements of an audiovisual product (in our experiment, recorded interviews) are repeated or rephrased by a professional transcriber to speech recognition software in order to generate a live transcript of the original dialogues/monologues (in our experiment, with a view to creating a translation using voice-over). This is similar to what in the USA is called (real-time) voice-writing/reporting and is used to produce verbatim reports in court, an area in which speech technologies have also been researched (Sohn 2004). This technique has also been used in the EPIC project to create corpus transcripts (Bendazzoli and Sandrelli 2005).

As for automatic speech recognition (ASR), also known as computer speech recognition or simply speech recognition, Anasuya and Katti (2009:181) present a review of the past sixty years of research on the topic and define it as "the process of converting a speech signal to a sequence of words, by means of an algorithm implemented as a computer program". In this case, the process is fully automatic with no dictation, but human revision is expected when certain quality thresholds are to be met. This is, in fact, the approach taken in our study.

2. Previous research on respeaking

Respeaking as a tool for the production of live subtitles has been the subject of increasing research in recent years, with Eugeni (2008a; 2009) and Romero-Fresco (2011) being two of the main authors in the field. Going back a few years, an excellent forum to gain an overview of incipient research on respeaking from the perspective of audiovisual translation was the International Seminar on New Technologies in Real Time Intralingual Subtitling, held in Forlì in 2006.**[1]** Most of the contributions focused on the process and professional practice of respeaking as implemented in different countries. The second conference,**[2]** held three years later in Barcelona focused on the reception of respoken subtitles and the strategies used by live subtitlers. The next conference in the series, held in Antwerp in 2011,**[3]** approached respeaking from the point of view of system developers, broadcasters and academics. The 2013 conference in Barcelona**[4]** featured the presentation of the Translectures project and the EU Bridge project, both of which looked into the automatic transcription and translation of online videos, lectures and webinars. Finally, the latest conference, held in Rome in 2015,**[5]** explored issues such as live subtitling quality on TV, respeaking by blind professionals and experimental studies on the reception of respeaking.

Beyond the scope of this conference series and others,**[6]** the need to train respeakers has generated several didactic proposals. These contributions often link the competences of respeakers to those of simultaneous interpreters (Eugeni 2008b; Arumí-Ribas and Romero-Fresco 2008; Romero-Fresco 2012a), given the multitasking and split-attention requirements of both disciplines.

The process of respeaking has also attracted the attention of researchers such as Van Waes, Leijten and Remael (2013), who have investigated the reduction strategies used by live subtitlers. Research on viewers' comprehension of respoken subtitles has also been conducted. Romero-Fresco (2010; 2012b) reports on the results of two experiments: the first aimed to test the comprehension of respoken subtitles for the news by hearing, hard-of-hearing and deaf viewers. Subtitles were presented at two different speeds and results show that most viewers find it difficult to obtain enough (verbal and visual) information from respoken subtitles with no sound. The second experiment was carried out using an eye-tracker and aimed to research how viewers read respoken word-

for-word subtitles as opposed to block subtitles. The results were significantly better for the latter than for the former.

More recently, research has been carried out on the quality of live subtitles, looking at the potential use of metrics. Traditional Word Error Rate (WER) methods have been adapted to the specificities of respeaking with models such as that developed by CRIM in Canada (Dumouchel *et al.* 2011) and the NER model (Romero-Fresco and Martínez 2015), which is being used by researchers, governmental regulators and broadcasters in different countries around the world (Ofcom 2013; MAA 2014). In 2013, the British regulator Ofcom commissioned a two-year research project to assess the quality of live subtitles on UK TV. This has resulted in the largest analysis of live subtitling available to date, providing data on the accuracy, speed, delay and edition rate of the live subtitles produced for 50 hours of live TV material, including 507,000 words and 78,000 subtitles (Romero-Fresco 2016).

Another leading project on respeaking is the EU-funded SAVAS project (2012-2014) (**http://cordis.europa.eu/fp7/ict/language-technologies/project-savas_en.html**), which aimed to develop speech recognition technology for live subtitling in different languages and in the domain of broadcast news, sports, interviews and debates. The project dealt with seven languages and created solutions for a speaker-independent transcription system, a collaborative respeaking-based system for live subtitling and the batch production of multilingual subtitles.

However, despite the interest raised by respeaking in the industry and amongst researchers, there is not much literature regarding the implementation of respeaking in areas other than TV, such as live events, or the comparison of its efficiency against ASR or manual transcription, which is currently the most common method used to transcribe non-fictional content. Sperber *et al.* (2013) propose a method for efficient off-line transcription of speech through respeaking via a combination of various techniques. Their method segments the speech input into short utterances, and selects only some of the utterances for respeaking based on confidence measure estimates. They present results from experiments by two respeakers and a simulation, and demonstrate the potential of their method in increasing speed and correction efficiency. However, the selection of respeaking, or typing as the better choice, is highly dependent on the particular segments. Bettinson (2013), on the other hand, investigates the application of respeaking to the transcription of data by field linguists, although for him transcribing means producing oral annotations (namely "careful speech" of the same content), a phrase level translation into a language of wider communication plus analytical comments. Also, for research purposes, Bendazzoli and Sandrelli (2007) use speech recognition software programs such as DNS or IBM Via Voice to obtain draft data which are later revised. The process explained in their study, which is referred to in the paper as shadowing because they adopt an interpreters' perspective, consists in listening to the recording and repeating aloud what the speaker says.

In the medical field, and mostly in the USA, researchers have looked into the feasibility of replacing manual transcription of patients' data with a semi-automatic approach based on respeaking (or voice writing, as this technique is known there). Al-Aynati *et al.* (2003) found transcriptions generated by dictation with speech recognition software to be a viable tool to transcribe pathology reports but also more time consuming than manual transcription, given the extra time needed to edit the errors caused by the speech recognition software. More recently, and taking advantage of more accurate software, Singh and Pal (2011) compared the feasibility and impact of making a transition from a manual transcriptionist-based service to one based on dictation with speech recognition in surgical pathology. The results of their experiment showed significant improvements in turnaround time and a positive impact on competency-based resident education. These findings have been echoed and confirmed in more recent studies (Joshi et al. 2014, Hartman 2015) which stress the usefulness of speech recognition-based transcription in the medical field.

Perhaps the most important study so far is the small-scale test conducted by D'Arcangelo and Cellini (2013) comparing the speed in the transcription of three samples of fast, medium-paced and slow speech using four methods: manual transcription, stenography, respeaking and automatic recognition plus correction. Although the test only offers anecdotal evidence, given that it compares the performance of two participants, it provides interesting data. Automatic recognition plus revision performs well in the transcription of slow speech and significantly worse when dealing with fast speech. As for the other methods, manual transcription is normally slower than respeaking and stenotyping, which is the fastest method.

Taking into account these encouraging results and the improvement in accuracy experienced by speech recognition software over the past years, the present article attempts to test speech recognition-based transcriptions in the film industry, which has so far relied extensively (and, to our knowledge, almost exclusively) on manual transcription (Pollak 2008).

3. Experimental set-up

The aim of the experiment presented in this article was to gather quantitative and qualitative data to compare three scenarios in the transcription of a non-fictional audiovisual excerpt: manual transcription, respeaking, and the revision (or post-editing) of a transcript generated by ASR. As stated in the introduction, our main interest was in the implementation of respeaking, hence our hypothesis was linked to this technique. A pilot test was carried out with five participants to validate the experimental set-up described in this section.

3.1. Participants

Ten native English participants (4 male, 6 female), all professional transcribers, took part in the experiment, which obtained ethical approval from the Universitat Autònoma de Barcelona. The quantitative data from two participants and the qualitative data from one participant could not be used due to technical reasons. All participants had professional experience in manual (QWERTY) transcription and none of them had used respeaking to transcribe audiovisual content before. Although an experiment with different profiles (for instance, professional respeakers versus professional transcribers) would have yielded interesting results, it was beyond the scope of this project and a homogeneous sample of participants was prioritized. In this regard, all of them were working as transcribers for independent companies that produce documentaries, reality TV shows, factual entertainment, as well as comedy and entertainment programmes. In terms of tools, they reported having used dedicated software such as Forscene, Aframe, Avid Interplay or Synergy but very often they simply used software to preview video files like VLC player and a word processing tool. Only one participant had used speech recognition software (DNS) to complete university assignments, but not for transcription purposes. The participants indicated different levels of satisfaction with their current jobs, 3.14 being the average on a 5-point Likert scale.

3.2. Materials

A 12-minute video interview was split into three four-minute excerpts, equivalent in terms of number of words. The video featured two female American Hip-Hop artists from California, Gavlyn and O'Blimey, as well as Kasia Ganzera, a young journalist working for an online portal called 'Hip-Hop Says'. The content of the interview was related to the recent work of the two artists, their European tour, their past experiences, and their future plans and aspirations. The language spoken was colloquial English, including many slang words (see annex for the transcriptions). This material was chosen by one of the authors of this study, himself a professional transcriber, as a representative example of material for transcription, featuring a real-life situation with unstructured, spontaneous colloquial speech for which no script is usually available.

Participants used Dragon Naturally Speaking 12 Premium to produce the respoken transcription of the video. In order to test the ASR method (automatic transcription plus manual revision), a comparative analysis was performed between the automatic transcription tool in Dragon and the state-of-the-art EML transcription server. The latter produced better results and was therefore chosen to generate the transcript for this part of the experiment. Even though the fact that this ASR system had not been trained specifically for the audiovisual content at hand may have an impact on the accuracy of the automatic transcription, it is an important requirement to test how a non-trained engine would perform with spontaneous speech.

Two pre-test questionnaires were prepared: one to gather demographic information at the beginning of the session and another one to find out about the participants' experience and views prior to the tasks. A post-test questionnaire was also prepared including questions on self-reported effort, boredom, confidence in the accuracy of the transcript and overall quality of each of the tasks (on a 5-point Likert scale), as well as subjective questions about the three transcription techniques considered here.

3.3. Procedure

The participants were welcomed in a computer lab and were informed about the general structure of the experiment. They filled in the pre-test demographic questionnaire and were provided with a 30-minute training session on respeaking by a professional respeaking trainer. After the session, they filled in the second pre-test questionnaire and proceeded to watch the video content. Each participant was then asked to transcribe three excerpts from the video interview using three methods: manual transcription, respeaking and revision of transcript generated by ASR. The order of the tasks and the videos used for each task was randomized and balanced across participants. In order to avoid fatigue, a maximum duration of 30 minutes was set per task, with a total of 90 minutes for the whole experiment. Those participants who managed to complete their transcription in 30 minutes or less were asked to note down the exact time spent on each task, whereas those who did not manage to complete their task noted down the exact time-code where they had stopped after 30 minutes. After completing the three tasks, the participants filled in the post-test questionnaire. Two researchers were in the room to monitor the participants and take observational notes. This was useful to discover, for instance, the mixed approach adopted by some of the participants who started to correct the ASR transcript, as discussed in section 4.

3.4. Methodology

Quantitative data was obtained on the time spent on each task as well as the ratio of minutes spent on the transcription per minute of original content. Qualitative data was also obtained via the questionnaires and the observation and interaction with participants. Descriptive statistics were used for the analysis.

The transcripts generated by all the participants were also assessed against a reference standard using an adapted version of the NER model (Romero-Fresco and Martínez 2015) and the NERStar tool (**http://www.speedchill.com/nerstar/**). The aim of this analysis was to detect the influence, if any, of the transcription method on the quality of the final output. The NER model defines accuracy with the following formula:

$$\text{Accuracy} = \frac{\text{N-E-R}}{\text{N}} \times 100$$

N refers to the number of words in the transcription (or subtitles), including commands and words. E refers to edition errors, generally caused by the subtitlers'/transcribers' decisions. R stands for recognition errors, meaning misrecognitions caused by mispronunciations or mishearings. E and R errors can be classified as serious, normal or minor, scoring 0.25, 0.5 and 1 respectively. For the purpose of this study and given that the NER model is designed for the analysis of live subtitles, the model was stripped of some of its subtitling-specific components, such as those that account for issues of delay, positioning and on-screen display. The different degrees of severity were still found to be valid and useful for the analysis of transcription as was the distinction between E and R. The latter accounted in this case for recognition and typing errors in the case of manual transcription.

4. Results and discussion

Regarding the time spent on the experiment, only three participants out of eight managed to complete the task in the required time (a maximum of 30 minutes) with manual transcription and with the revision of the ASR transcript, whilst in respeaking the number increased to five participants out of eight.

Considering only the data of those participants who finished (three from manual and ASR revision, five from respeaking), manual transcription is the fastest technique and the revision of the ASR transcript is the slowest one (see Table 1). As for the quality of the output as calculated with the NER metrics, the highest quality was obtained with the revision of the ASR transcript, followed by manual transcription and respeaking. It was observed that some participants started off correcting the automatically-generated transcript but then deleted the remaining content and transcribed it manually. For this reason, figures about the revision of the ASR script are often closer to those of manual transcription. Future research with higher quality ASR (or with less spontaneous material) may yield different results, but this behavior may provide interesting insight into the transcribers'

attitude towards ASR depending on its accuracy rate, all of which could be analysed with keylogging software. Respeaking has the lowest NER values, which may be due to the fact that most participants did not perform a thorough revision of the respoken text. Instead, they just provided the raw respoken results. In any case, the NER score in all the conditions is relatively low if compared, for instance, with the minimum score for subtitles to be considered as acceptable, which is 98% (Ofcom 2015). Yet, as mentioned in the introduction, while live subtitles are produced by fully qualified respeakers, using a trained speech recognition software to generate an output that is used by billions of viewers, the objective of transcription is usually different, as different are the conditions in which this experiment was carried out. It is for this reason that 98% may be seen as an excessively high threshold for transcription, and further research on metrics for this type of transcription technique should be carried out.

	ASR	Respeaking	Manual
Speed (time spent on one minute of programme)	6'54''	6'26''	5'18''
Accuracy (NER)	98.02	96.87	97.7

Table 1: Mean data from participants who finished the task

When considering all the participants (see Table 2), manual transcription was again the fastest option, followed by respeaking and the revision of the ASR transcript. Once again, some participants felt frustrated with the revision of the automatic output, deleted all the content and manually transcribed the excerpt, which is a direct consequence of our experimental design and should be taken into account in future research.

Concerning NER values, manual transcription yields the highest scores, followed by the revision of the ASR transcript and respeaking. None of them, however, reaches the above-mentioned 98% threshold.

	ASR	Respeaking	Manual
Speed (time spent on one minute of programme)	9'36''	8'36''	7'39''
Accuracy (NER)	97.535	97.161	97.783

Table 2: Mean data from all participants

These data are in line with previous results and show that, overall, manual transcription is the fastest option. This is a logical consequence of the participants' experience as professional manual transcribers who have developed their own techniques. Accuracy in manual transcription is the highest when considering all the participants and the second highest when considering only those who completed the task. Once again, their professional practice with manual transcription generally involves revision. What is, however, surprising, is that in manual transcriptions their accuracy is below 98%, which in other fields such as subtitling has been established as the minimum NER score to provide a quality output. This shows again that research is needed on what the minimum score for a quality transcription should be.

Respeaking was the second fastest in the experiment. It is the method that allowed the highest number of participants to finish, but it also slowed down other participants, thus showing a wide range of speeds. It obtained the lowest results in terms of accuracy, mainly because most participants did not revise the respoken transcript. These results show the potential of this technique for the transcription of audiovisual material but also two aspects that must be factored in the equation: training and time for revision. The participants had a 30-minute crash course on the basics of respeaking, while the average training for respeakers ranges from 3 weeks to 3 months (Romero-Fresco 2012a). Future research will also need to account for the impact of revision on the overall speed of transcription, which may be higher with respeaking and ASR than with manual transcription.

As for the revision of an ASR transcript, the results are very much determined by the approach adopted by the different participants. Most of them began by attempting to correct the ASR transcript, but then decided to delete it and change to manual transcription. It then became a type of mixed approach, a slower version of manual transcription with higher accuracy because revision was at the heart of it. However, the increase in time did not result in an increase in quality in all cases. Although quality was higher than with manual transcription and respeaking when considering only the participants who completed the task in 30 minutes, it was lower than with manual transcription when considering all the participants

Regarding qualitative and more subjective data, Table 3 includes a comparison between the participants' opinion on ASR and respeaking before the experiment (with no knowledge or experience on ASR and respeaking other than the 30-minute crash course they had just received) and after the experiment. Table 3 includes the statements with which they had to indicate their level of agreement on a 5-point Likert scale, next to the pre-task and post-task mean data.

Statement	**Pre-task mean**	**Post-task mean**
Manual transcribing is too time consuming	3.4	3.2
Respeaking could be a useful tool to transcribe documentaries	4.5	3.8
Automatic speech recognition could be a useful tool to transcribe documentaries.	4.1	2.7
Respeaking could speed up the process of transcription	4.5	3.9
Automatic speech recognition could speed up the process of transcription	4.1	2.1
Respeaking could increase the accuracy of transcriptions	3.8	2.9
Automatic speech recognition could increase the accuracy of	3.0	2.2

transcriptions		
Respeaking could increase the overall quality of transcriptions	3.4	3.1
Automatic speech recognition could increase the overall quality of transcriptions.	2.8	2.5

Table 3: Pre-task and post-task opinions (mean data on a 5-point Likert scale)

Before the experiment, participants expected respeaking to be a very useful technique to transcribe documentaries (4.5) and to increase the speed (4.5), accuracy (3.8) and overall quality (3.4) of the transcriptions. These expectations dropped to different extents after the experiment, with 3.8 for usefulness, 3.9 for speed, 2.9 for accuracy and 3.1 for overall quality. In other words, having completed the experiment with a mere 30-minute training session on respeaking, most participants regard this method as useful, fast (especially given that they consider manual transcription, which they do for a living, too time-consuming) and with an acceptable quality. They do have doubts, however, regarding the accuracy of the transcriptions.

In contrast, the expectations raised by ASR were not met after the experiment. On paper, ASR looked useful (4.1) and it was expected to speed up the process (4.1), the accuracy (3.00) and, to a lesser extent, the overall quality (2.8) of the transcriptions. After the experiment, ASR was no longer regarded as useful (2.7), fast (2.1), accurate (2.2) and of good quality (2.5), with all items under analysis below 2.5 on the scale. Interestingly enough, though, the NER metrics show that quality of the ASR transcription did increase, which indicates that it may be the self-reported effort (see Table 4) that makes the participants think that the accuracy of ASR is so low.

After the experiment, participants were also asked about their perceptions in terms of effort, boredom, accuracy and overall quality with regard to the transcripts they generated. Results are summarised in Table 4.

	Respeaking	**ASR**	**Manual**
Self-reported effort	2.89	4.55	3.11
Boredom	2.22	3.89	3.12
Accuracy	2.78	2.89	4.22
Overall quality	3.22	3.00	4.33

Table 4: Participants' assessment

Respeaking is regarded as the least taxing (2.89) and least boring (2.22) method, in contrast particularly with the revision of an ASR transcript (4.55 and 3.89, respectively) but also with manual transcription (3.11 and 3.12, respectively). In other words, respeaking seems like a very attractive method with which the participants would like to replace manual transcription. However, they still harbour doubts regarding its overall quality (3.22) and especially its accuracy (2.78), which is reasonable given the objective results shown in tables 1 and 2. In these two aspects, manual transcription is still regarded as more reliable (4.22 and 4.33, respectively), even though objective metrics show that in reality the revision of the ASR transcript was more accurate than the manual one.

A final open question asking participants about their general views on the experiment and the possible use of respeaking for the transcription of audiovisual material provided some interesting additional insights, which we summarise next.

Participant 1 was "impressed with the re-speaking software... especially when used in combination with manual input. The video material was using slang language and I feel that re-speaking could work even better in an interview situation with (standard) English and no slang". As indicated before, this excerpt was chosen because it was particularly challenging and because it is a stereotypical example of extemporaneous audiovisual content. However, as suggested by this participant, other types of content could probably yield even better results. This statement also points indirectly to an interesting area of research, that of automatically assessing what content is worth respeaking and what content should be manually transcribed. In addition, the observation of the participants in the experiment showed what may be the most effective transcription method: a combination of manual transcription and respeaking. Instead of opting for hands-free respeaking, some participants chose to combine respeaking with the occasional use of typing, mostly to correct errors, thus making the most of the speed inherent to respeaking and of their proficient typing skills. Striking the right balance between manual transcription and respeaking could be a good solution to increase the speed and (crucially) the accuracy of transcriptions. This can only be achieved after proper training and practice, none of which could be provided in the experimental setting. Some participants suggested in informal post-test interviews a combination of 75% voice and 25% manual transcription as an ideal balance, but this remains to be verified by further research.

Participant 4 thought respeaking could be useful to improve accuracy for those transcribers who cannot spell very well, whilst stressing that more specific training was required. Participant 5 also highlighted the need for more practice, as did Participant 6, who added some interesting insight concerning his/her job as a manual transcriber: "I am a fast typist but I am used to a different keyboard, so this slowed me down on the manual transcription. I am also used to using a transcription software such as Express Scribe, which makes manual transcription a bit faster. I was a lot slower with Dragon because it takes time to learn how to use the different options such as go back, vocab, etc. I think in the long term this would be much quicker than manual transcription."

On the other hand, most views on ASR are negative, such as those expressed by Participant 2, who explains that s/he had to re-do everything by hand, or Participant 3, who considers that the technology still needs to improve. It is important to point out that the ASR system was not adapted to the domain or the speakers in the video, which undoubtedly affected the quality of the results. A better performance could be expected when using an engine trained on the specific domain of the audiovisual material tested.

To conclude, the participants were asked to answer the question "Would you enjoy your job more than you enjoy it now if you used respeaking for transcription?" Five possible replies were given on a 5-point Likert scale, and results show that 22.22% would enjoy it much more and 66.66% would enjoy it a bit more, while only 11.11%

consider they would have the same level of enjoyment and none considers they would enjoy it less than manual transcription.

5. Conclusions

This article has presented a first exploratory study on the potential use of respeaking as a tool for transcribing non-fictional audiovisual material through an experiment in which three conditions were compared, namely manual transcription, respeaking and the revision of transcript generated by ASR. Although limited in its number of participants, our research has provided both quantitative and qualitative data that will hopefully be useful for, and expanded in, future research.

Manual transcription was the fastest option, followed by respeaking and ASR, but it was respeaking that allowed the highest number of participants to finish the transcription in the time allocated for the task. The downside is that accuracy levels in all conditions were unexpectedly low (and lower than the 98% accuracy rate required in other fields such as subtitling), especially in respeaking. On the one hand, this makes us think that further research is needed to ascertain what the minimum NER score for a quality transcription should be, a value that may be linked to the purposes of such transcription. On the other, this shows that manual transcription is still the most effective method in terms of speed and accuracy, which comes as no surprise taking into account that participants were professional manual transcribers. Still, as shown by subjective opinions gathered via questionnaires, transcribers are fairly tired of the manual method. They regard it as too time-consuming, taxing and boring, and are willing to try out new possibilities. Therefore, the results of the experiment only partially fulfill the initial hypothesis, which suggested that respeaking could be successfully used to speed up the process of transcription and could have a positive impact on the transcribers' experience.

When presented with alternative methods, they initially embrace the idea of using ASR, but after the test they find it is less useful than expected. In our experiment, the revision of a script generated by ASR does not reduce transcription time because many participants end up doing a manual transcription from scratch after deleting the ASR content. However, further research should be conducted with this method, since the conditions used in this experiment (an ASR engine that was not adapted to the genre of the audiovisual material tested) did not allow the system to reach its standard performance rates.

Respeaking is very welcome by participants and performs very well in terms of speed, even though participants were not familiar with this technique and the training was extremely short and not suited to the specificities of this type of transcription. As indicated by one of the participants after the experiment, respeaking was perceived as a very fast system which could solve spelling issues but, in order to make it accurate, specific training and a clear method would be needed. The training provided was short and focused on producing subtitles, whilst for professional transcribers the specificities of the task at hand should be taken into account. As observed in informal talks after the experiment, it seems that a combination of respeaking and manual transcription could be the solution.

From a research perspective, it would be worth investigating a method with a quality control system to automatically propose to the end user the most efficient transcription method for a given video, including ASR if it proves to be reliable enough. It would also make sense to conduct a more specific study comparing the speed and accuracy of manual transcription versus respeaking taking into account not only the transcription time but also the revision time. Another aspect worth investigating is the effect of longer working sessions, since the results for this experiment were obtained in tasks limited to a maximum of thirty minutes. It remains to be seen how professionals would perceive respeaking in terms of effort and boredom after longer sessions. Experiments with a higher number of participants and different audiovisual material are definitely needed to yield more conclusive results. For instance, in voiced-over non-fictional genres translators usually edit out many features such as repetitions, false starts, hesitations and non-relevant discourse markers. Working with an edited text generated by respeaking could maybe speed up the process or even facilitate machine translation plus post-editing, an aspect worth researching.

Many research possibilities emerge and further experimental studies that can have a clear impact on the industry will hopefully be derived from this first experiment. For the time being, transcribers seem ready to test respeaking in order to make their jobs not only more efficient but also more enjoyable.

Acknowledgments

This article is part of the project ALST (reference code FFI-2012-31024), funded by the Spanish Ministerio de Economía y Competitividad. Anna Matamala and Pablo Romero-Fresco are members of TransMedia Catalonia, a research group funded by the Catalan Government (reference code 2014SGR0027). We would like to thank all the professional transcribers who took part in the experiment. Special thanks to Juan Martínez and the European Media Laboratory (EML GmbH) for their support.

Appendix: Transcription Excerpts

Clip 1

00:01 L: okay I'm rolling

K: yeah, okay, let's go, hey, what's up guys today we are with Gavlyn and O'Blimey

O: it's good

G: hey

00:12

K: guys, so, crazy thing that I saw you 2 years ago in the cipher effect, right, that was like the first one and then I'm like, okay, that's cool, dope stuff, then you're featured on O'Blimey's ill video and on the whole track, and now you are on your European tour

O, G: <laughs>

K: guys, that's crazy, how, how like, how did that happen, like, how comes it took you like two years to kind of do something together?

G: well she's lived in San Francisco for a minute, she just barely moved to Los Angeles and the first time we met was at the cipher effects, so

K: oh, was it? Okay, alright

O: I just moved down to LA

00:50

O: and was like living in LA half time, and San Francisco half time, working on my music, so, so really like making time to collaborate with a bunch of people in LA, just was a happening, you know at the time and as soon as I was there full-time, you know I let Gavlyn know that I was back in town and ready to work as well as the other girls from the cipher too, we've all collaborated a bunch, but Gav, I was like I have a song for you right now, she called me on her way back from another Europe tour

K: okay

01:18

O: and she was like, you want to get together? You want to work on something? I was like, I have a song and studio time ready to go right now, so she came into the studio that night, right after getting back from the airport and we recorded 'ill' and literally since that song we have hang out like every other day, ever since that point, so

G: yeah

K: okay, so it was cool for you to actually spend so much time together, because I mean European tour, you've got like, I don't know 30 shows or something like that

G: we've got 20

O: 25

G: 25

O: yeah

K: that's crazy long time, but, like how do you feel about this whole tour, because it must be a lot of work and travelling and you guys kind of working together as well, like how does that make you feel

01:55

G: it feels awesome, like, I consider O'Blimey one of my best friends so it's super tight to be touring with my girl and kill it and see the people enjoy what we do together, it's an amazing feeling, I love it

O: like, no doubt that there's tonnes of work, but it's kind of like getting to go on holiday or going on vacation, you know

G: yeah

O: you know with your best buds, we may annoy the shit out of Mike Steez our DJ and manager from time to time, but I think the three of us together, we have a good team, and

K: it's good

O2: 21

O: it's just, we vibe, we mesh, you know, it's easy

K: yeah, so, what would you say is the best thing about hip-hop culture, like what's the thing that you love the most about it?

G: you start

O: the best thing about hip-hop culture? Man, I mean, a very, you know, a very wise person once said is that you know the best thing about music, and it goes for hip-hop for me, is that when it hits you you feel no pain

K: yeah

02:47

O: and you know, whether or not a track, or a vibe in an area brings you to a painful place, the fact that everybody else, you know, kind of understands and has been there as well, whether or not the context matches perfectly, but just the fact that other people understand what you are going through and there is an entire community that goes through it as well

K: that is a great thought

G: exactly

O: especially like, with the honesty in hip-hop right now, everybody is just telling it like it is and I think that, it's rare, it's beautiful

03:15

G: I have to second on what she is saying, like, that it's the coolest thing about hip-hop, it's like just being able to write shit and do shit that, you feeling explains who you are, you know what I mean? And people are enjoying that and vibing out with it, it's super tight

K: cool, okay, I saw your hip-hop Kemp the other time and I was like, okay, crazy show you and Organised Threat on the show and Yella Brother was with you as well, I was like, okay, you literally just just burnt the stage down, then I'm on my way to RA the Rugged Man, I was like yeah, waiting for the show

03:50

K: and then I see you guys on the stage as well, I was like, what is happening? Like, how did you make to be on the stage with the guy and how does that feel? I mean he is a great artist, I mean crazy, but great

Clip 2

00:01

G: I mean shout out to R A the Rugged Man, he's become a really good friend, what's it called, no, it was cool, I mean, we're, we, like, we hit it off and we were chopping it with RA, Yella Brother knew who RA was and at that time RA broke his ankle, so he was

K: yeah, I saw him on the wheelchair

G: yeah, he was rapping on a wheelchair and he was like yo, I need someone to back me up, y'all down to back me up? And we were like alright and that's literally how it happened and then we went into a car and practiced the song for 20 minutes and he said alright we are up and I was like holy shit, okay, we are doing this

K: it was great show though

G: thank you

00:32

K: I mean, the vibe and all that plus you know, just him being on the stage it makes you feel like, you don't know what's gonna to happen the next minute, at least you know the songs, you know, show is like a different world especially with that guy

G: she did throw a few sticks in the crowd, so it's crazy

K: yeah, and I've got a question for you as well, cause I know that you have background with like jazz music and blues music in your family, so you sing as well, which is kind of, well a rare thing for an MC to do, so how did hip-hop came about then?

01:05

O: oh man, that's a great question, you did your research

K: I did yeah

O: You're killing the game, you know, growing up obviously my dad was a blues musician, so I grew up on the side of stages and you know, I think that blues and hip-hop are a lot alike, you know, it gives you really a space to talk about hardships and struggles and experiences, so when I was a teenager, or preteen, like 12-13 years old, you know, I needed an outlet that was more relevant to me, you know, and on the schoolyard everybody was freestyle battling each other and stuff like this, so I went from writing poems and songs on my guitar to being like, I think I want to step in the ring and, you know, in the year of 8th grade, like 12 years old, I just made the switch, I just put down my guitar and I said I'm going to step into the ring and start battling

01:58

O: only recently have I really started to dig deeper into the singing and stuff like that, because I really took like 10 years away from singing and, my girlfriend actually back home, she is a beautiful singer, she vocally produced my whole last album, she has actually been encouraging me to sing a bit more and stuff like that, I feel way more comfortable rapping, but the music I listen to the most is like R&B, you know and base and oldies and jazz, like, I just want to vibe out, like really heavy, so, she's like if you love it so much just start writing it

02:31

O: and so I have, it's really cool, it's the sensitive side

K: second confirmation, yeah, because, I mean, you must feel that it is a bit different when there is a girl on the stage and when there is a guy in the stage, obviously like, we know, how it is, like hip-hop, like it doesn't matter, like your age, race, nothing really, but I guess it might be so, do you get a sort of like a different vibration from people, like, do people feel like, uhm I'm not really sure how is that going to go?

02:58

O: totally, low expectations, well, now that, I mean obviously with Gavlyn, you know she's blown the fuck-up, especially, you know seeing her on stage, people already know what to expect from her, so it's different for me to be on this tour, where people know what to expect from a girl that's really good at rapping, but you know, playing for a crowd that doesn't, you know has no idea what's gonna come, you know it does come with low expectations, and when you can break those expectations and really just blow it out of the water, and make people surprised, I think that's one of the greatest feelings and accomplishments, so, (K: yeah) despite those low expectations, like, it's a challenge kind of, we have the opportunity to show somebody

G: yeah

O: something against the stereotype

K: girl power

O: exactly

03:43

G: like two girls on stage are already, like oh God

O: what are they doing?

K: and then it's like jaw dropping, oh like my God what's happening?

G: it always happens every time

03:52

K: that's good

G: so it's exciting

K: like, imagine that you would have all the money, all the talents in the world, would you still do what you

Clip 3

00:01

K: do, would you, would you want to try something else, something new? Did you ever thought about that?

G: I would probably do what I do and mesh it with a bunch of shit

K: okay

G if I was like, had all the extra talents that I wanted, I would be like cool I'm gonna to turn this rap music into something

00:20

K: so it would be like dancing, and singing and the DJ'ing at the same time

G: all of that, all of that, yep

K: okay, what about you O'Blimey?

O: it's crazy like, I feel like this has been a common question I've had in my head it's like, what would I do with all the funding in the world, you know

K: yeah

O: because, starting as like an independent artist, you know, not really having the money to start off with, you really see how vital it is to kind of make a budget and delegate, you know certain funds to certain areas, but thinking about what if I started with a bunch of money, where would I be right now, it's like crazy, you know, I could be DJ'ing, you know and I could have the voice box on stage, so I could be singing like crazy songs and like, you know, I would also probably like run a gym and have you know crazy fitness like programmes going on and stuff, you know, because I'm really into that sort of stuff, but like, just like the little sacrifices that you have to make as an independent artist running an independent career and when you think about not having to make those financial sacrifices, and everything just kind of being financially more easy, it's a difficult question to ask yourself, you know, where would I be if? But,

01:25

K: well, it's hard to imagine, so it's like, you know one way you lived, and then you know, the thing is that I think people also think about what they could have done, but actually people do have the money or something, or already are up there, they just think like like you people as well, because obviously, you do what you love, and they do what they do as well

O: totally

K: like, if there is money or connections or whatever, yeah,

O: I get it

K: I guess it's something that you can think about, it's good you guys would still do what you do which means like you are really dedicated, pretty dope actually, would you name like three things that you can't live without, whether it is like food, drink, music, I don't know people, places, anything, like three things

02:10

O: Wi-Fi

K: yes, that would be my one probably

G: music's definitely one of them too

O: girls, for me, personally

G: <laughs>

02:31

G: sure, okay I would say music, food and weed <laughs>

O: music food and weed <laughs>

02:41

O: so Wi-Fi, music, girls

K: alright, that's a nice combination as well, guys, so is there anything that we should be waiting for right now, do you have any more, obviously materials, you've got some video clips coming up, just shout out what you've got coming up

02:59

G: sure, I'm working on my brand-new project called make up for your breakup, it's like, it's probably the most mature sexual, feminine record I've ever done

O: it's super cool, it's crazy good

G: and I feel it's my best shit ever and other than that I do have a video clip coming out this thing called black chain cool lady, keep your eyes peeled for that and already have 2 records out: 'from the art' and 'modest confidence' go get that at Gavlyn.com

03:28

O: I am, I just finished a mix tape, shout out to Luke, who executive produced that, also executive producing her new mix tape which is gonna be incredible, I finished a mix tape called manifesto, which is m-n-f-s-t-o and there's gonna be a couple of new music videos coming out for that and a couple singles that I have got coming out, a song called Petaluma, which I, which actually is the only singing track that I have and you guys got to hear a little bit, so a video for Petaluma and a single for Petaluma coming out and yeah

04:01

K: that's great, make sure you guys check them out, so thanks again Gavlyn and O'Blimey

G: thank you for having us

O: shout out to say what

G: yeah, shout out say what

K: okay

References

Al-Aynati, Maamoun and Chorneyko, Katherine A. (2003) “Comparison of voice-automated transcription and human transcription in generating pathology reports”, *Archives of pathology and laboratory medicine* 127, no 6: 721-725.

Anasuya, M.A. and Katti, S.K. (2009) “Speech recognition by machine: A review”, *International Journal of Computer Science and Information Security* 6, no 3:181-205.

Arumí-Ribas, Marta and Romero-Fresco, Pablo (2008) “A practical proposal for the training of respeakers”, *Journal of Specialised Translation* 10:106-127.

Bailey, Julia (2008) "First steps in qualitative data analysis: transcribing", *Family Practice* 25, no 2: 127-131.

Bendazzoli, Claudio and Sandrelli, Annalisa (2007) “An approach to corpus-based interpreting studies: developing EPIC (European Parliament Interpreting Corpus” in *Proceedings of the Marie Curie Euroconferences MuTra: Challenges of Multidimensional Translation*, Sandra Nauert (ed), Saarbrücken, available online: **http://www.euroconferences.info/proceedings/2005_Proceedings/2005_Bendazzoli_Sandrelli.pdf** (last accessed 18 October 2016)

Bettinson, Mat (2013) *The effect of respeaking on transcription accuracy*. Honours thesis, unpublished. Melbourne: University of Melbourne.

Bucholtz, Mary (2000) “The politics of transcription”, *Journal of Pragmatics* 32: 1439–1465.

Cole, Alaistair (2015) *Good Morning. Grade One. Language ideologies and multilingualism within primary education in rural Zambia*, PhD diss, University of Edinburgh.

Davidson, Christina (2009) “Transcriptions: imperatives for qualitative research”, *International Journal of Qualitative Methods,* 8, no 2: 35-62.

Díaz-Cintas, Jorge and Orero, Pilar (2006) “Screen translation: voice-over” in *Encyclopedia of languages*, Keith Brown (ed), London: Elsevier, 473.

D’Arcangelo, Rossella and Cellini, Francesco (2013) “Metodi e tempi di una verbalizzazione - prove tecniche” in *Numero monografico sul Respeaking, Specializzazione on-line*, Carlo Eugeni and Luigi Zambelli (eds), 81-95.

Dumouchel, Pierre, Boulianne, Gilles and Brousseau, Julie (2011) “Measures for quality of closed captioning” in *Audiovisual translation in close-up*, Adriana Serban, Anna Matamala and Jean-Marc Lavour (eds), Bern, Peter Lang: 161-172.

Eugeni, Carlo (2003) “Respeaking the BBC News. A strategic analysis of Respeaking on the BBC”, *The sign language translator and interpreter,* 3, no 1: 29-68.

Eugeni, Carlo (2008a) *La sottotitolazione in diretta TV. Analisi strategica del rispeakeraggio verbatim di BBC News*, PhD diss, Università Degli Studi di Napoli Federico II.

Eugeni, Carlo (2008b) “A sociolinguistic approach to real-time subtitling: respeaking vs. shadowing and simultaneous interpreting” in *English in international deaf communication*, Cynthia J. Kellet Bidoli and Elana Ochse (eds), Bern, Peter Lang: 357-382.

Eugeni, Carlo and Mack, Gabriele (eds) (2006) New technologies in real time intralingual subtitling. *Intralinea. Special issue: Respeaking.*

Franco, Eliana, Matamala, Anna and Orero, Pilar (2000) *Voice-over translation: an overview*, Bern, Peter Lang.

Green, Judith, Franquiz, Maria, and Dixon, Carol (1997) “The myth of the objective transcript: Transcribing as a situated act”, *TESOL Quarterly*, 21, no 1: 172–176.

Hartman, Douglas J. (2015) “Enhancing and Customizing Laboratory Information Systems to Improve/Enhance Pathologist Workflow”, *Surgical Pathology Clinics*, 8, no 2: 137-43.

Joshi, Vivek, Narra, Vamsi, Joshi, Kailash, Lee, Kyootai, and Melson, David (2014) “PACS administrators' and radiologists' perspective on the importance of features for PACS selection”, *Journal of Digital Imaging*, 27, no 4: 486-95.

Lambourne, Andrew (2006) “Subtitle respeaking”, in *New technologies in real time intralingual subtitling. Intralinea. Special issue*, Carlo Eugeni and Gabriele Mack (eds), **http://www.intralinea.org/specials/article/1686** (last accessed 01 March 2016)

MAA (Media Access Australia) (2014) *Caption quality: international approaches to standards and measurement*, Sydney, Media Access Australia **http://www.mediaaccess.org.au/research-policy/white-papers/caption-quality-international-approaches-to-standards-and-measurement** (last accessed 01 March 2016)

Mereghetti, Emiliano (2006) “La necessità dei sordi”, in *New technologies in real time intralingual subtitling. Intralinea. Special issue*, Carlo Eugeni and Gabriele Mack (eds), **http://www.intralinea.org/specials/article/1697** (last accessed 01 March 2016)

Muzii, Luigi (2006) “Respeaking e localizzazione”, in *New technologies in real time intralingual subtitling. Intralinea. Special issue*, Carlo Eugeni and Gabriele Mack (eds), **http://www.intralinea.org/specials/article/1688** (last accessed 01 March 2016)

Ochs, Elinor (1979) “Transcription as theory”, in *Developmental pragmatics*, Elionor Ochs and Bambi Schiefflin (eds), New York: Academic, 43-72.

Ofcom (Office of Communications) (2013) *Measuring the quality of live subtitling: Statement*, London, Ofcom. **http://stakeholders.ofcom.org.uk/binaries/consultations/subtitling/statement/qos-statement.pdf** (last accessed 01 March 2016)

Oliver, Daniel G., Serovich, Julianne M., and Mason, Tina L. (2005) “Constraints and opportunities with interview transcription: Towards reflection in qualitative research”, *Social Forces*, 84, no 2: 1273–1289.

Pollak, Alexander (2008) “Analyzing TV Documentaries”, in *Qualitative Discourse Analysis in the Social Sciences*, Ruth Wodak and Micha Krzyanowski (eds), Houndmills: Basingstoke, Hampshire, 77-95.

Prazac, Ales, Loose, Zdenek, Trmal, Jan, *et al.* (2012) “Novel approach to live captioning through re-speaking: tailoring speech recognition to re-speaker’s needs”, *InterSpeech 2012:* 1372-1375.

Ramondelli, Fausto (2006) “La sottotitolazione in diretta con la stenotipia” in *New technologies in real time intralingual subtitling. Intralinea. Special issue*, Carlo Eugeni and Gabriele Mack (eds), **http://www.intralinea.org/specials/article/1694** (last accessed 01 March 2016)

Romero-Fresco, Pablo (2010) “Standing on quicksand: viewers’ comprehension and reading patterns of respoken subtitles for the news”, in *New insights into audiovisual translation and media accessibility*, Jorge Díaz-Cintas, Anna Matamala and Josélia Neves (eds), Amsterdam, Rodopi, 175-195.

Romero-Fresco, Pablo (2011) *Subtitling through speech recognition: Respeaking*, Manchester, Routledge.

Romero-Fresco, Pablo (2012a) Respeaking in translator training curricula: present and future prospects, *The Interpreter and Translator Trainer* 6, no 1:91-112.

Romero-Fresco, Pablo (2012b) “Quality in live subtitling: the reception of respoken subtitles in the UK”, in *Audiovisual translation and media accessibility at the crossroads,* Aline Remael, Pilar Orero and Mary Carroll (eds), Amsterdam, Rodopi, 111-133.

Romero-Fresco, Pablo (2013) “Accessible filmmaking: Joining the dots between audiovisual translation, accessibility and filmmaking”, *Jostrans - The Journal of Specialised Translation* 20: 201–223.

Romero-Fresco, Pablo (2016) “Accessing communication: The quality of live subtitles in the UK”, *Language & Communication* 49: 56–69.

Romero-Fresco, Pablo and Martínez, Juan (2015) “Accuracy rate in live subtitling – the NER model”, in *Audiovisual translation in a global context. Mapping an ever-changing landscape,* Rocío Baños Piñeiro, and Jorge Díaz-Cintas (eds), London: Palgrave Macmillan, 28-50.

Singh, Meenakshi, Pal, Timothy (2011) “Voice recognition technology implementation in surgical pathology: advantages and limitations*”, Archives of Pathology and Laboratory Medicine*, 135, no 11: 1476-81.

Sohn, Shara D. (2004) *Court reporting: can it keep up with technology or will it be replaced by voice recognition or electronic recording?* Honours thesis, Southern Illinois University. **http://opensiuc.lib.siu.edu/cgi/viewcontent.cgi?article=1264&context=uhp_theses** (last accessed 01 March 2016)

Sperber, Matthias, Neubig, Graham, Fügen, Christian, *et al.* (2013) “Efficient speech transcription through respeaking”, *InterSpeech 2013:* 1087-1091.

Trivulzio, Fausto (2006) “Natura non facit saltus”, in *New technologies in real time intralingual subtitling. Intralinea. Special issue*., Carlo Eugeni and Gabriele Mack (eds), **http://www.intralinea.org/specials/article/1690** (last accessed 01 March 2016)

Van Waes, Luuk, Leijten, Mariëlle and Remael, Aline (2013) “Live subtitling with speech recognition. Causes and consequences of text reduction”, *Across languages and cultures,* 14, no 1:15-46.

Wald, Mike (2009) “Real-time speech recognition subtitling in education”, *2nd International Seminar on Real-Time Intralingual Subtitling.* Barcelona, UAB.

Notes

[1] http://www.intralinea.org/specials/respeaking

[2] http://www.respeaking.net/barcelona_2009.html

[3] http://www.respeaking.net/antwerp%202011.html

[4] http://www.respeaking.net/barcelona%202013.html

[5] http://www.unint.eu/it/component/content/article/8-pagina/494-respeaking-live-subtitling-and-accessibility.html

[6] See, for instance, the 49th Intersteno Congress (**http://www.intersteno2013.org/**) and the recent Let’s Talk, organised by the Australian regulator ACMA (**http://www.acma.gov.au/theACMA/live-captioning-lets-talk-register**),

---

"The Use of Respeaking for the Transcription of Non-Fictional Genres An Exploratory Study", *inTRAlinea* Vol. 19.
Stable URL: http://www.intralinea.org/archive/article/2262