Fostering access for all through respeaking at live events

Zoe Moores, University of Roehampton

ABSTRACT

Live subtitling using speech recognition, known as respeaking, is widely used to make live television programmes accessible. Although a growing area within audiovisual translation internationally, in the UK its industry use has been limited to television, in part due to the many misconceptions surrounding its production. This study explores how respeaking can be introduced to complement current access provision at unscripted or partially scripted events. Through close collaboration with users and providers, respeaking is shown to be a viable way of providing access for deaf, deafened and hard of hearing audience members in this new sector: access that a wider audience is also likely to benefit from. The paper begins with a brief discussion of the audiovisual landscape, focusing on quality in respeaking and current provision in the sector. Next, a bespoke training programme is presented and user and provider expectations for the service are outlined. Finally, the quality of respeaking at a series of research events is discussed. The results indicate that the quality attained for the most part meets, and frequently exceeds, the benchmark of 98% accuracy set for live television subtitling. Latency is similar to that seen on television, yet remains an area for further consideration.

KEYWORDS

Respeaking, live subtitling, accessibility, d/Deaf, deafened, hard of hearing, professional training, quality, NER, action research.

1. Introduction

Introduced to UK television in the early 2000s, respeaking is a mode of creating subtitles through speech recognition (Lambourne 2006; Romero-Fresco 2011, 2018b) that provides sensorial access for people who are d/Deaf, deafened and hard of hearing (DH) to live programmes, including the news, sports and special events. The fact that these subtitles are created in real-time adds particular challenges to the production process, associated with the display rate of the subtitles, latency, technical issues (frozen, lost or badly-positioned subtitles) and errors. As the limits of speech recognition are pushed due to the fast-paced speech seen on television, errors, and sometimes unusual errors at that, are an undeniable feature of respeaking. Despite the complexities of the production process, expectations for live subtitles are high: as with traditional pre-recorded subtitles for deaf and hard of hearing people (SDH), viewers expect full access to the programme being broadcast so the quality of the subtitles, both in terms of grammatical accuracy and coverage of content is under close scrutiny (Ofcom 2013a, 2013b). The two-year study led by Ofcom into the quality of live subtitles on television demonstrated that, in addition to ensuring that subtitling quotas set by Ofcom (2017) are met, the quality seen in respoken subtitles has continued to improve (Romero-Fresco 2016). Respeaking is also a growing field within audiovisual translation and media accessibility. It is used in many different countries (Romero-Fresco 2018b)

and current pan-European research projects into the application of respeaking in interlingual settings[1] and into the professionalisation of the industry[2] demonstrate its growing and future potential.

In many countries, there is already a tradition of using respeaking to provide access in live settings outside of the television sector. For example, in Flanders, both intralingual and interlingual respeaking are being introduced at conferences; in Poland, respeaking has been used at meetings of the Polish parliament. In Germany, Austria and Switzerland, speech recognition software is used in combination with a specialised set of shortcuts that are programmed on regular keyboards. Referred to as speech-to-text interpreting (STTI), this service provides access in educational, political, social and medical settings (Eichmeyer 2017). In the USA, voice writers, the term used for respeakers, have joined stenographers, who traditionally provided Communication Access Realtime Translation (CART) or Live Event Captioning at conferences, events, classes and discussions (CCAC 2016a, 2016b).

In the UK, respeaking has barely made an entrance into the live event sector (see section 2.2). Instead, it remains a method of subtitle production that is little understood and frequently criticised. The increasing presence of respoken subtitles on public screens in airports, pubs and waiting rooms means that most people, and not only those who choose to view them on individual screens, have been exposed to them. Nevertheless, few understand how they are actually produced and the general perception is that these subtitles have been typed, and typed badly at that. Whilst respeaking is undoubtedly a profession, perhaps current lack of professional status for respeakers (Romero-Fresco *et al*. forthcoming), together with these public perceptions, mean that attempts at introducing respeaking outside television have had limited success.

Yet, there is a need for increased sensorial access outside of television so that everyone can benefit from it. Currently, excellent sensorial access does exist at live events, as will be outlined in section 2.2 below. However, without the regulation of an equivalent body to Ofcom, the quantity of the access provided in the cultural sector does not compare to that seen on television. The Equality Act 2010 (Legislation.gov.uk 2010: online) states that "reasonable adjustments" to services are expected. Whilst we understand that these adjustments may relate to physical features, auxiliary aids and/or the provision of information (Section 20), the details of exactly what these entail remain vague. The European Accessibility Act 2015 and directive on the accessibility requirements for products and services which followed in 2019, Directive (EU) 201/882 (EUR-Lex 2019) specify a range of devices and technologies which are to be made accessible, and include many which might be used in the context of live events, for example, when booking tickets or accessing content on mobile devices. However, as with the Equality Act, precise requirements for the cultural sector are not set. The question posed here is whether respeaking

could be introduced into the live event sector to increase the amount of access provided. If so, what would need to be in place, in terms of training and workplace procedures, to ensure that the quality of this access matched industry standards and audience expectations?

The aim of this paper is to answer these questions by analysing the access provided across a series of research events. The paper begins with a brief discussion of the live audiovisual landscape, giving an overview of quality in respeaking, the importance of inclusive access and current access provision in the live sector. Next, content of the initial training given to the respeakers to prepare them for this new setting is revealed, along with the user and provider expectations that this service is designed to meet. Finally, an industry-standard, user-centred, analysis of the respeaking produced leads to a discussion of the quality of respeaking at live events and enables the pathway for equipping respeaking professionals from the regulated environment of a television Access Services department to work in diverse live scenarios to be illustrated.

2. Live audiovisual landscape

2.1 What is respeaking?

Respeaking is the production of subtitles in real time by speech recognition. There are three key stages within respeaking. First, the respeaker listens to the broadcast content and speaks the aural content of the programme, voicing in punctuation, sound labels and any additional content that needs to appear in the subtitle. As they do this, they may edit the original spoken content slightly, perhaps adjusting the order or punctuation of the original, or omitting or paraphrasing certain words (Eugeni 2006; Lambourne 2006; Romero-Fresco 2011). Next, the speech recognition software processes the input. Finally, the recognised utterances pass through the subtitling software and the respeaker is able to make further, slight, adjustments to the subtitles as or after they appear on screen (McIntyre *et al*. 2018). Even while the second and third stages are happening, the respeaker must continue with step one, as the audio of the broadcast continues, determining the speed of respeaking required. Where sections of a programme are scripted and the respeaker has access to them in advance, for example in the news, the respeaker is also able to cue subtitles out as-live, rather than voicing them in.

For this reason, respeaking can be considered as a form of "computer-aided simultaneous interpretation" (Romero-Fresco 2018b: 96), with two key differences from 'pure' interpretation: firstly, as is the case here, respeaking is usually intralingual and, secondly, the intermediary of speech recognition software demands that the words spoken conform to the capabilities and limitations of the speech recognition tool in use, rather than the human ear. A carefully prepared voice model and good audio are essential for

respeaking. Further, if accurate subtitles are to be conveyed, intense concentration is required.

As a form of live subtitling, respeaking can be used to create subtitles at high speed, which reach high levels of accuracy (Romero-Fresco 2011). With initial training lasting two to three months, the costs involved in setting up respeaking are on the low side, in particular when compared to traditional stenotyping. This is a key reason that respeaking has spread so widely on television. As with any form of live interpretation, there is an inherent delay or latency in respeaking between a word being spoken and appearing on screen in a subtitle; this results from the time needed for the spoken word to be heard, respoken, recognised and processed through the subtitling software and onto the screen (McIntyre *et al*. 2018). Usually this delay is low, although technical issues can mean spikes are seen. Ofcom's (2015) guidelines suggest there should be a maximum latency of 3 seconds. During the final report on the quality of live subtitling in 2015, an average latency of 5.6 seconds was seen across the range of programmes sampled, whilst the average across the rounds was 5.3 seconds (*ibid*.). Where there was no as-live cueing, latency averaged 7-8 seconds, with peaks of 10-21 seconds possible.

2.2 Access at live events

Is respeaking viable for the live event setting? To answer this requires a closer consideration of the term 'live event'. In this context, 'live' is used to refer to an event happening in real-time, where the audience attends in person, and which is not watched in its entirety through a screen, although parts of it (the subtitles and certain visual elements such as PowerPoint slides or video clips) will necessarily be displayed on one (or more)[3]. Live-subtitled events on television therefore fall outside the scope of this study, as they are being watched in their entirety through a screen; if, however, additional respeaking were to be provided for the audience present at the recording, then that respeaking would fall within the definition of a live event used here.

It is also important to note that the use of respeaking is only being suggested at unscripted or partially scripted events. Where a full script exists, such as at the theatre or opera, preparing captions, surtitles or supratitles[4] in advance and cueing them out live is the better option for access (Díaz Cintas and Remael 2007; Stagetext 2011; Mele Scorcia 2018). However, where the event is for the most part unscripted, for example Q&As and discussions, or, perhaps, semi-scripted, but the exact words spoken may vary, for example at presentations, talks and tours, the suggestion is that respeaking could be used.

The inherent variety present within the potential content and staging of so-called live events also means that the use of the term 'subtitle' to describe the respoken content must be questioned. Since this content can be

displayed almost anywhere around the main event, as will be illustrated below, the locational specification implied within the 'sub' of subtitle no longer applies. For this reason, going forward, I will use the term 'titles' as a more accurate term to refer to the respoken content at live events. On occasion, I will adopt the term (sub)title, when reference is made to both subtitles on television and the respoken titles at live events.

What access is currently provided at live events in the UK? Stagetext (stagetext.org) is a national charity that advocates for and provides access to theatre shows and live events for people who are DH through accessible text. In 2000, when theatre captioning first began in the UK, there were nine captioned performances; by 2009/2010, there were 200 captioned performances in venues around the UK. The following year, Stagetext began working with museums, galleries and cultural venues to make talks and lectures accessible through live titles[5], known as speech-to-text transcription (STT), produced using special shorthand keyboards. In 2010/2011, Stagetext provided access at 188 captioned performances and 16 live subtitled events. In 2017/2018, the total of captioned performances had risen to 351 and live subtitled events had risen to 238[6].

Through their captions and live titles, Stagetext aims to provide access for the diverse audience that experience hearing loss, access that does not depend on an understanding of sign language interpretation; however, sign language interpreters are sometimes present at their live talks and tours to accommodate the diverse audience who attend them. Many events are also made accessible solely through sign language interpretation.

There is no single record for the percentage of live events that are accessible for people who are DH across the UK. However, when we consider the number of venues around the country and the number of events taking place on a daily basis, the information suggests that the proportion of accessible events is low. For example, the State of Museum Access Report (Cock *et al.* 2018), which audits museums considered to be the best in the UK, reveals that only 3% mentioned BSL interpreted talks on their websites and only 1% referred to titled talks or tours. Whilst the actual provision may vary, as the report states, "access and inclusion starts online" (*ibid.*: 5). Potential visitors use the information about access that is available online to decide whether or not to visit in person. By failing to mention accessible services that are on offer, future visitors may be lost.

The purpose of this research is therefore to explore how the number of events that are accessible can be increased and how respeaking can play a role in this, by complementing the access provision which is already in place. Providing this access is a fundamental issue not only of equality, but also of equity. Language is a fundamental tool for being able to participate and, if a person cannot access the language, what is expressed in that language also becomes inaccessible. While people who are DH are the original target audience for this access, the wider audience who may also

benefit from these titles is considered early on in the process (Udo and Fels 2010; Romero-Fresco 2018a), in line with the Universalist take on access (Greco 2018). From the outset, consideration is given to what needs to be done to make venues and events accessible on a number of levels, including physically, sensorially, linguistically and socially. Thought of in this way, access becomes far more than the traditional question of mobility and how to enter a venue (Cock *et al.* 2018: 19). Just as language is a tool for participation, access itself becomes a tool to enable full participation and engagement in what is taking place. Equality means that everyone receives access while equity ensures that the form it comes in makes it fit for purpose (Mann 2014) and quality is central to this.

2.3 Quality assessment in respeaking

For the access provided by respeaking to be effective, it has to be reliable and of a good quality. The NER model (Romero-Fresco and Martínez 2015) is the baseline for Ofcom's 2013-2015 review of quality in live subtitling and provides an assessment that is "research-informed, valid, reliable and user-focused" (Romero-Fresco forthcoming). The original and respoken transcripts are compared, word for word, and both edition (E) and recognition (R) errors are weighted and scored according to the impact they have on audience reception (see section 4.5), then deducted from the total number of words spoken by the respeaker (N); from this, an accuracy score is calculated (McIntyre *et al.* 2018). A context-based comment provides a more descriptive and summative account of the quality of the respeaking and the model takes into account the effects of speed and delay in its scoring. The reliability of the model is demonstrated by the high interrater consistency seen of 0.09% (Romero-Fresco 2016). Within this model, 98% accuracy is considered acceptable and a score above 99.5% is considered excellent (nertrial.com n.d.). Whilst the threshold of 98% might seem high, the weighted scoring system means that subtitles that achieve this could still contain certain errors that have a serious impact on the viewer's comprehension, numerous minor errors, or a combination of both (Ofcom 2014). Further, recent research in Poland and Canada has found positive correlations between the NER scores and users' views of the quality of live subtitling output, indicating the validity of this threshold (Romero-Fresco 2016, forthcoming; Szarkowska *et al*. 2018; CRTC 2019). The average accuracy for the four rounds of sampling, across all channels and genres, was 98.38%. In the final round of the Ofcom (2015) study, an average of 98.55% accuracy was seen.

While SDH on television now provides an impressive model for other access services in terms of both quantity and quality, in the beginning the focus was on quantity of subtitle coverage alone (Romero-Fresco forthcoming). When moving to the live event setting, a more appropriate place to focus is quality: by replicating the respeaking quality already demonstrated on television, reassurance can be given to critics that this is a viable model, and from there, as attitudes change, it is hoped that quantity will grow.

However, in order to ensure that the quality of respeaking seen on television could be transferred into the live event setting, it became clear that respeakers would need specific and effective training, which would take into account the new and variable environment of live events. This was a central feature of the research design.

3. Conception of the study: Focus groups and respeaker training

In order to determine with certainty whether respeaking should be introduced into the live event sector, an assessment of the quality of respeaking at live events is needed, not simply at a single event, but in a way that is proven to be reliable and that can be replicated. Using the NER model would ensure that the assessment is user-focused and would enable comparisons to be made with established quality standards within TV subtitling. However, providing and watching titles at a live event differs greatly from doing this on television. The opportunity was therefore taken to work closely with a number of focus groups so that user expectations could be understood and the service could be mapped in a way which allowed others to follow. A structure which interwove focus groups and action research was chosen for this purpose, as shown in Figure 1[7]:

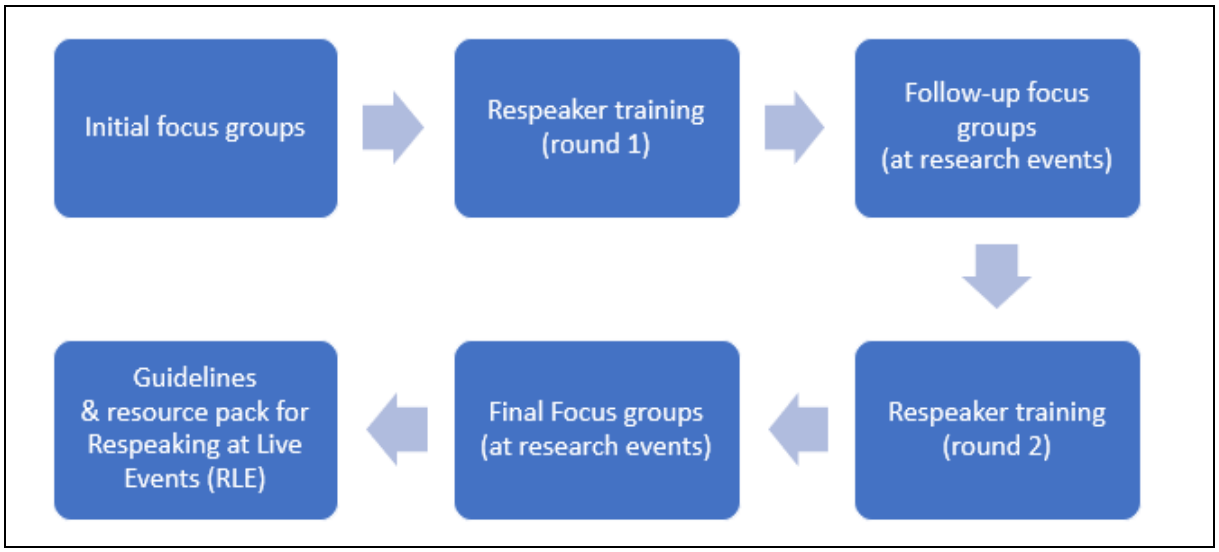

Figure 1. Research Cycles

Respeaking at live events was initially targeted at DH audience members as it provides sensorial access to events and this focus group had the largest reach. However, just as television subtitles are used by a wider audience base, it was thought that other members of the audience would benefit from this provision. For this reason, audience interviews were also held with hearing non-native speakers of English (NNE) as there is evidence to suggest that intralingual subtitles can facilitate and support language learning (Vanderplank 2016).

Fifteen DH people attended the initial focus groups, with six people taking part in the first and nine in the second. Their hearing levels were mixed and

they ranged in age from mid-twenties to over 60. Those attending had a mixture of occupations and many were now retired. The call for volunteers had been published on the Stagetext website and shared at the University of Roehampton and with people who had previously participated in related research projects there. Almost everyone who attended was already familiar with subtitles on television so, for that reason, these focus groups were used to discuss a range of features within subtitles and how they could be transferred to the live setting.

The points raised during these group sessions were incorporated into a wider survey, completed in full by 55 people, where participants were asked to rank their priorities for live subtitling on television and at live events. Respondents were provided with a list of fourteen criteria to rank with a Likert-style scale, from 'not at all important' to 'extremely important'. They were also able to add new categories, within their own questionnaire, which they felt should be considered and rank them on the same scale. The priority ranking was very similar for both scenarios: having little or no delay, an easy to read font, specialist terminology subtitled correctly, subtitling that reflected either the full content or very near to the full content, little or no obstruction of the picture and error-free subtitles were key. Requests were also made for indications within the subtitles when a foreign language was being spoken, when content was omitted and when something was inaudible.

The interviews with the NNE were used to explore how useful they found the intralingual (sub)titles. The sample size was very small and consisted of seven people, aged between 29 and 67. Their native languages included Dutch, Korean and Luganda. Employment information was not collected, but education ranged from college to Masters level. Even with such a small sample, some interesting insights were offered into how intralingual (sub)titles could be beneficial for this group and this is certainly an area worthy of further exploration. Most participants agreed that the (sub)titles helped them understand and engage with accents, faster speech and lyrics; they also found seeing the spelling of particular words important. When asked to rank their priorities for (sub)titles, having little or no obstruction of the picture, (sub)titling key words and well-timed (sub)titles were key. Whilst the motivation behind the priorities of the two audience groups differed (DH participants citing a desire for equal access and NNE prioritizing the visual image), there was consensus in the expectations they set for live titling. Further, since these applied equally to television and live events, this seemed to confirm that experiencing both settings would be beneficial for respeakers in practical and professional terms.

In order to consider how to set up the service, two further focus groups consisted of the service providers, i.e. the venues that host events and the respeakers themselves. Interviews and visits were carried out at five venues. Twelve in-person interviews were held with respeakers and a further six respeakers replied to digital questionnaires; these respeakers

came from three companies that provide access services. For both the venues and respeakers, ensuring a reliable service that met the demands of users was key. In particular, early interviews with respeakers revealed that a working set-up which prioritised the respeakers' ability to focus and gave them access to good quality audio would be essential; they also wanted clarity on what was expected of them in this new setting, both in terms of the access they would be providing, but also in terms of how their role might evolve when present in person at events. Once the research events began, the speakers and presenters that took part formed an additional group, who straddled the divide of user and provider. Their views on the use and implementation of respeaking at live events were also sought (see section 4.3).

The shift from working in the known entity of TV access services into unknown locations where events may be held constitutes a significant change for a respeaker. Equipment, set-up, audience and the content being titled, the keystones of their regular work, are all subject to change. Devising a training programme that prepared them for these challenges was essential. Action research was the adopted approach as it allowed respeaking to be tested across a range of scenarios and incorporated stages of action and reflection. It allowed the respeakers to become part of a "participative community of inquiry" (Reason and Bradbury 2008: 1). The content covered in training and decisions made at the research events were critically examined and evaluated by both the respeakers and myself as researcher. Learning from one event was incorporated in the next and procedures evolved from event to event, as well as from cycle to cycle, ensuring the final training programme was a robust model, applicable to diverse settings. The role of the respeakers in this process was central, acting as they did as informed participants.

Four live subtitlers participated in this process. All were experienced television respeakers and had worked in the profession for between three and seven years at the time of training. None had any experience of respeaking at live events. Since the minimum period needed for basic training as a respeaker is two to three months (Romero-Fresco 2011), and many respeakers cite one and a half to two years as the timeframe needed to pass a confidence threshold in their work, as researcher, I was confident in the ability of the participants and the quality of their respeaking.

The respeakers participated in fourteen hours of initial training, split across three sessions and six modules. Five modules were researcher-led and provided the respeakers with an overview of the project and the expectations which existed for respeakers and respeaking at live events, and introduced them to the equipment being used and the practicalities of setting up in new environments; respeaking style and conventions were also discussed. The final module was split across numerous sessions and allowed the respeakers to practice and familiarise themselves with the new equipment and set-up; it included personalised and tailored training and

provided the respeakers with an opportunity to create a new voice model, since they were unable to use their professional profiles during the research events.

The research events were the first opportunity for these respeakers to provide access at a live event and to meet and receive direct feedback from the audience.

4. The study

In order to determine whether respeaking can be introduced successfully into the live event setting, an analysis is presented here of respeaking at a selection of research events. The respeaking is analysed with the NER model, within which, allowing for addition and recognition errors, the respoken content should reach a minimum of 98% accuracy. In line with action research methodology, each event provided an opportunity for learning and refinement to the use of respeaking at live events; whilst this was not a specific criterion for quality, these opportunities nevertheless contributed to attaining quality and will therefore be mapped through this discussion. An overview of all eight research events will be provided; NER data is shared for seven of the events, since the recordings made at the fifth event were not of sufficient quality to permit a NER analysis. Two events are analysed in detail as case studies to demonstrate this process in action.

4.1 Event design

A total of eight research events were held over the course of a year. The first four events ran in autumn 2017 and the second four in the summer of 2018, allowing a period for analysis, reflection and further training between each round. The events lasted between one and two hours each and were held in various locations around the UK.

The scope of the events was broad and they included presentations, public speakers, film panels, museum tours and post-screening Q&As, as illustrated in Figure 2:

	Event type	Venue	Respeakers	Software	Working set-up	Audiovisual content
Round One						
1	Live arts presentation (seated)	Riverhouse Barn, Walton on Thames	A, B	Text on Top	Respeakers located at back of theatre (raised) Performers on flat stage Titles at top of screen	Multiple speakers Live performance PowerPoint with text and images
2	Public speaker (seated)	Riverhouse Barn, Walton on Thames	C, D	Text on Top	Respeakers located at back of theatre Presenter on raised stage Titles at bottom of screen	Single speaker PowerPoint with text and images
3	Film panel (seated)	BFI, St Stephen Street, London	A, C	Text on Top	Respeakers in separate room Presenters on flat stage Titles at top of screen	Multiple speakers Some use of PowerPoint with text and images Videos with burnt in subtitles
4	Museum tour (mobile)	Wellcome Collection, London	B, D	Streamtext	Presenters located within museum with a mobile phone audio stream from the presenter Titles on individual tablets	Spoken word and visual realia
Round Two						
5	Post-screening discussion (seated)	The Watershed, Bristol	A, C	Text on Top	Main discussion group located in open café Respeakers located at the table next to the discussion group Multiple lines of titles streaming to two different screens	None during the discussion
6	Post-screening discussion (seated)	The Depot, Lewes	C, D	Text on Top	Respeakers located at side of screening room Presenters sat on flat stage Three lines of titles in the middle of the screen	None during the discussion
7	Highlights museum tour (mobile)	Manchester Art Gallery	A, D	Streamtext	Respeakers located within the museum with a mobile phone audio stream from the presenter Titles on individual tablets	Art works around the museum
8	Public speaker (seated)	University of Roehampton	B, C	Text on Top	Respeakers located at the back of the lecture theatre (raised) Presenter on flat stage Three lines of titles in the middle of the screen	No slides; use of gestures and action to support content

Figure 2. Event Overview

This design allowed respeaking to be tested across single/multiple speakers, a seated/moving audience, diverse technical set-ups in and outside the event room and varied visual and spoken subject matter. Variation in the venue, event type and content were planned in advance; the variations in the working set-up were determined on site according to the specifics of the location. In a similar set-up to television work, the respeakers worked in

pairs to provide the access at each event. Whenever individual schedules allowed, the pairings were alternated to facilitate discussion and reflection across events.

In the initial research design, the intention was for the audience to experience the feeling of attending a 'real' live event rather than a research event and it was hoped that the research element of the event would be as invisible or unobtrusive as possible. In the first round of events, this was difficult to achieve as the explorative and trial nature of the research was so evident and two events (one and three) were created specifically to provide an opportunity for testing. In the second round of events, access was added to existing events and the research element was less obtrusive, though undeniably present. The events in the first round were based around London/Surrey to enable participants to attend multiple events. In order to achieve wider geographical reach, three events out of the final four were held further afield in Bristol, Lewes and Manchester.

4.2 Equipment

Each respeaker had a laptop, USB mouse and keyboard. The speech recognition software was Dragon Individual Professional and the subtitling software was Text on Top at seated events (Fig. 3), and Streamtext at the mobile events, as this allowed streaming onto multiple screens via the internet (Fig. 4):

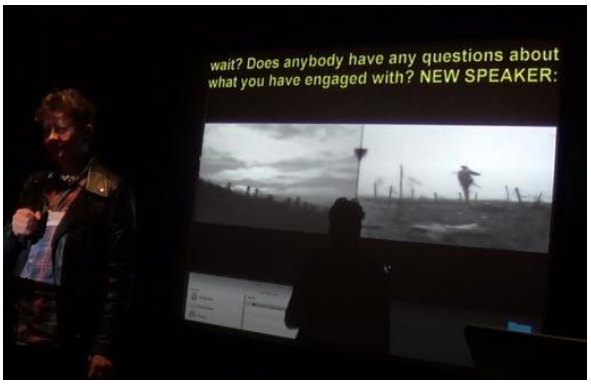

Figure 3. Text on Top software in use at the BFI (event 3)[8]

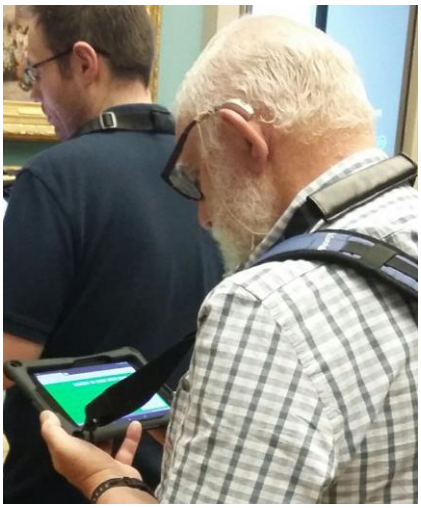

Figure 4. Titles streaming to individual tablets via Streamtext at Manchester Art Gallery (event 7)

Sylencer face mask microphones were purchased from Talk Technologies as they muffled the respeakers' voices, allowing them to work in the main event room when required. Various mouthpieces were available for the microphone. Some only covered the mouth and some covered the mouth and nose. The respeakers each chose the one they were most comfortable with. A special strap was used to hold the microphone in place. Talk

Technologies have since released a microphone stand and this will be offered to respeakers at future events. Silent Disco headsets enabled the respeakers to receive good quality audio directly from the speakers' microphones. An LED light was also purchased so the respeakers could work in darker environments. These items made up the respeakers' individual kits and could be fitted within a rucksack for portability (Fig. 5):

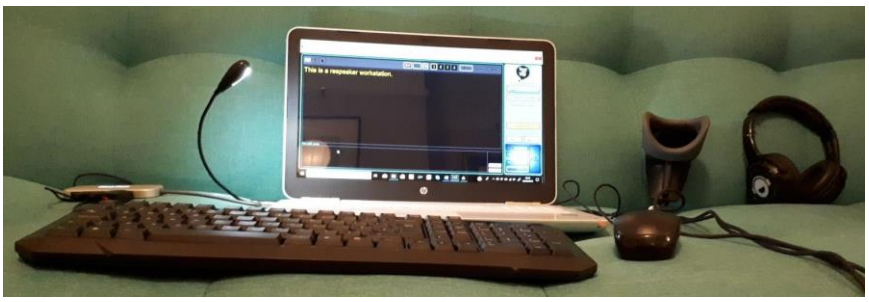

Figure 5: Individual kits

A dual wireless handheld microphone system was chosen for the flexibility it provided the speakers and a Falcon 3-channel transmitter sent the signal to the headsets. In round one, VH2 microphones from QTX were purchased and used with either the Citronic 2-channel compact mixer or the venues' own systems. This changing set-up proved complex and problematic and it quickly became clear that the success of the event depended on the quality of the microphones. For this reason, a complete Respeaking at Live Events (RLE) kit was created, consisting of the Falcon transmitter, the Shure SM56 wireless microphone combination system, and a Xenyx X1222USB mixer and microphone stand. A TX208 Alto speaker allowed amplification of voices in larger venues.

4.3 Participants

On average, 14 people attended each event. They had mixed hearing levels and a range of native languages, including BSL, and a variety of spoken languages. Their ages ranged from 13 to 88. The events were advertised through Stagetext, Action on Hearing Loss (AOHL), the National Association for Deafened People (NADP), on social media and through word of mouth. Sixty five people attended a single event, and 16 people attended more than one, including one person who attended all eight.

The presenters varied at each event. At events one and three, they were experienced live artists (event one) and film producers (event three), who were also personal contacts. Public speakers (events two and eight) were found on the Diane Mannering website and the tour guides (events four and seven) regularly led tours at the museums in question. Events five and six were discussion based, so presentation was shared between the chair and the audience alike. As researcher, I also presented during sections of each event.

4.4 Procedure and experimental design

Before each event, I conducted a site visit for familiarisation with the venue. Decisions were taken about the location of the respeakers, arrangement of the room and title display position. In venues where some in-house equipment was being used, a technical check was run to ensure compatibility with the respeakers' microphones, headsets and subtitling software. In the museums, the internet signal was checked along the route of the tour. Information about respeaking and guidelines for presenters were shared with presenters and key venue staff. The presenters were asked to outline the content of the talk (key themes and terminology) and to list any technical requirements and visual aids being used. This information was shared with the respeakers around a week before the event.

On the day of the event, I arrived on site with the respeakers a few hours in advance of the audience. One hour was scheduled for them to train in new vocabulary and to review live event conventions, such as handovers. Technical checks were also scheduled with the venue staff. There was an opportunity for the respeakers and presenters to meet before the event began so that the presenters could learn more about respeaking and experience being respoken and the respeakers could raise any queries about the content of the event and familiarise themselves with the speech and manner of the presenters.

I began each event with a brief introduction, where the format of the event and the questionnaire that audience members would be asked to complete were both explained. Audience members were also invited to participate in a focus group discussion at the end of the event. Following this briefing, the presenters ran the event as they usually would, and at the end of the event, once the questionnaires had been completed by hand, the focus group began. BSL interpretation was provided during the introduction and focus group; the entire event was respoken.

At the end of the event, the respeakers, presenters and key venue staff were also invited to share their experience of the event in a short questionnaire. Paper copies of the questionnaires were provided, and the option of completing a digital version of the same questionnaire was also offered.

In order to analyse the respeaking, video and audio recordings were made at each event and the transcripts of the respeakers' output were saved. At the seated events, cameras were positioned to record the whole event, while at the mobile events, a camera followed one tablet throughout the tour to record the titles that appeared.

4.5 NER analysis: overview

In line with the methods adopted by Ofcom when measuring the quality of live subtitling on television (Romero-Fresco 2016), a sample of 10-16 minutes from each event was analysed using the NER model.

Care was taken to ensure that the samples were representative of the different events. Figure 2 above lists the different types of audio (single/multiple speakers) and the visual content (slides, gestures, videos) present at each event. This was used as a basic checklist of features to include when selecting the clip, to which examples of audience interaction were added. In line with the action research approach, I recorded my own observations at each event, noting particular features and complexities, and sought to include these within the samples. For most events, it was possible to select a single stretch of recording which included all these features. At two events, composite clips from two shorter passages were created.

Trends seen within the data collected at these events were compared with those observed in the Ofcom study, which is the largest corpus of respoken subtitles in the UK, consisting of 78,000 subtitles and 546,000 words. A further comparison was also made with the LiRICS corpus (see endnote 2), which consists of 6,000 subtitles and 40,000 words. Both corpora are in English.

In the NER analysis, recognition and edition errors are classified as serious, standard or minor, with penalties deducted accordingly. Serious errors result in the audience being misled and carry a weighting of 1. Standard errors carry a weighting of 0.5. In the case of recognition errors, these are the ones the audience is likely to notice and be confused by while a standard edition error usually happens when an independent idea is omitted. Minor errors are weighted less at 0.25. Minor recognition errors can be easily understood by the audience and minor edition errors mean dependent details are omitted (Romero-Fresco and Martínez 2015). The totals shown in Figures 6 and 7 below indicate the total deductions made for each error category. Instances where respeakers correct an error are listed as 'corrections' and, in these cases, no penalty is taken for the original error. In addition to editions where content is lost, respeakers also make many correct editions, where words may be omitted or paraphrased; in these instances, there is no loss of information for the viewer and there is no penalty for these (McIntyre *et al*. 2018).

	Round One				Round Two		
Event	**1**	**2**	**3**	**4**	**6**	**7**	**8**
Edition Errors:							
Serious	1	1	3	2	2	3	1
Standard	11	7	17	20	20	11	7
Minor	24	17	19	22	22	12	23
Total	36	25	39	44	44	26	31
Recognition Errors:							
Serious	0	1	0	2	1	1	0
Standard	0	1	0	3	0	6	1
Minor	71	107	111	43	113	16	112
Total	71	109	111	48	114	23	113
Accuracy	98.2%	97.3%	98.1%	97.8%	97.3%	98.7%	98.5%
Rating	Acceptable	Substandard	Acceptable	Substandard	Substandard	Good	Good

Figure 6. Accuracy data (including software-specific errors)

	Round One				Round Two		
Event	**1**	**2**	**3**	**4**	**6**	**7**	**8**
Edition Errors:							
Serious	1	1	3	2	2	3	1
Standard	11	7	17	20	20	11	7
Minor	24	17	19	22	22	13	23
Total	36	25	39	44	44	27	31
Recognition Errors:							
Serious	0	1	0	2	1	0	1
Standard	0	1	0	0	0	0	1
Minor (excluding software errors)	9	27	17	23	41	15	18
Total (excluding software errors)	9	29	17	25	42	15	20
Revised accuracy*	99.1%	98.8%	98.9%	98.3%	98.4%	99.0%	99.1%
Revised Rating*	Very good	Good	Good	Acceptable	Acceptable	Very Good	Very good

Figure 7. Revised accuracy data* (excluding software-specific errors)

I had initially expected to be able to use the NER model exactly as it was conceived for intralingual respeaking on television to analyse the respeaking at live events. However, as the analysis proceeded, it became clear that certain error types appeared within the live event setting that stretched the regular pathways of analysis present within the NER model. Whilst the error severity outlined above did not change, the process involved in identifying the errors was more complex; discussion of some of these errors follows

below. At the end of the RLE study, an adapted version of the NER model for Live Event analysis was therefore proposed.

Common practice when assessing with the NER model is for a clip to be marked by two independent reviewers, allowing an interrater agreement to be calculated. The process here differed slightly as the NER model was adapted during this process. As researcher, I completed the first marking for all the events. An external reviewer then completed sample-second marking, reviewing approximately 30 titles per event. Following this, and incorporating the second marker's comments, I second marked the events in full, whilst adapting the NER model for live events, and then completed a full review for consistency. Marking comments reveal instances of different scoring, but no interrater agreement was calculated.

Before commenting on the spread of errors, one further note is required about software-specific errors (Moores and Romero-Fresco 2015), since these led to two different scores for accuracy being calculated at the live events. Software-specific errors occur when an unintended transcription is produced, despite the respeaker speaking correctly. Different forms of software-specific errors occur, but all are classified as recognition errors. Some relate specifically to Dragon's vocabulary: for example, when a respeaker utters a word that is not in the vocabulary, alternative words will appear. Sometimes a contextually incorrect spelling may appear for a word that is in Dragon's vocabulary. These errors are often preventable with good preparation. Errors may also occur which the respeaker cannot pre-empt in their preparation: sometimes respoken content fails to appear due to transmission-related problems and sometimes errors occur that are due either to the (sub)titling software or to its (lack of) compatibility with Dragon.

The errors that occurred in high numbers fell into the latter categories. At the mobile events, there were moments when the respoken content was not transmitted in full to the tablets being used by the audience, since the internet connection temporarily dropped. At many events, duplicated spaces between words (by far the most common), missing spaces between words and instances of doubled initial letters, as illustrated in Figure 8, were also seen, which seemed to result from the interaction between Dragon and the titling software.

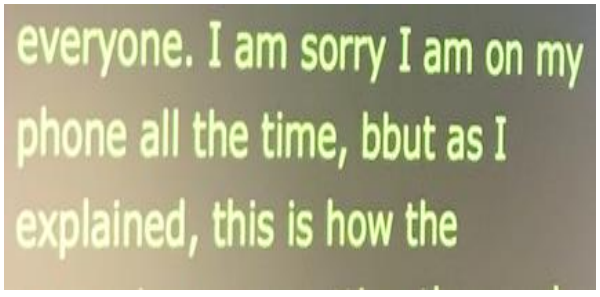

Figure 8. Doubled initial letter

The first accuracy score includes every error in the count (Fig. 6) and reflects the titles seen by the audience. Whilst there will always be glitches and errors at live events, it seemed that these particular errors both masked the quality of the respeaking that took place and were ones that could be overcome with software updates and improved internet connections. Having an accuracy score that excluded them was therefore felt to be of great research interest. A revised accuracy score (Fig. 7) was thus calculated to determine what the accuracy would have been without these particular software-specific errors. Figure 9 compares these two data sets:

	Round One				Round Two		
Event	**1**	**2**	**3**	**4**	**6**	**7**	**8**
Accuracy	98.2%	97.3%	98.1%	97.8%	97.3%	98.7%	98.5%
Rating	Acceptable	Substandard	Acceptable	Substandard	Substandard	Good	Good
Revised accuracy*	99.1%	98.8%	98.9%	98.3%	98.4%	99.0%	99.1%
Revised Rating*	Very good	Good	Good	Acceptable	Acceptable	Very Good	Very good

Figure 9. Accuracy data compared

The accuracy rates seen are as might be expected for professional, experienced respeakers transferring to a new setting. They range from substandard to good, when software errors are included, and from good to very good, when excluded. The highest accuracy scores, both including and excluding errors, were seen in the final two events, where the respeakers had grown more accustomed to this new setting. However, the revised score at the first event was comparably high and variation across event type is also visible (see section 4.6). This illustrates that to fully understand the NER score given, the accuracy percentage must be interpreted alongside the comment that accompanies it. Here, details of the event context and any errors particular to the live event setting are noted and analysed.

Many similarities can be seen when the data collected at live events is compared with the Ofcom and LiRICS corpora. As seen in Figure 10, the average accuracy rate at live events, including errors, is 98.0% (acceptable), slightly lower than that seen in the Ofcom and LiRICS corpora. When software errors are excluded, the average rating rises to 98.8%, an average that is higher than that seen in Ofcom and LiRICS and which narrowly misses a 'very good' rating.

	RLE	RLE* (excluding software errors)	Ofcom	LiRICS
Average accuracy rate	98.0% (Acceptable)	98.9% (Good)	98.4% (Acceptable)	98.5% (Good)
Excellent (sub)titles	0%	0%	5%	7%
Very good (sub)titles	0%	43%	22%	30%
Good (sub)titles	29%	29%	29%	33%
Acceptable (sub)titles	29%	29%	21%	7%
Substandard (sub)titles	43%	0%	23%	23%

Figure 10. Summary of results for the RLE, the RLE*, Ofcom and LiRICS pilot

As the Ofcom data in Figure 11 illustrates, the common trend amongst professional respeakers is for there to be a higher proportion of edition errors than recognition errors. One key reason for this is the challenge that keeping up with the original audio poses. At live events, where respeakers will be working in unfamiliar settings, a similar trend could be expected.

	RLE	RLE* (excluding software)	Ofcom	LiRICS
Total edition errors	29%	61%	69%	65%
Total recognition errors	71%	39%	31%	35%

Figure 11. Total edition and recognition errors

Once the software-specific errors are excluded from the RLE data, the same trend is seen and it seems that the respeakers' attention is, primarily, going towards editing. The number of recognition errors in the RLE corpus is higher than that in Ofcom and LiRICS. This can be attributed to a number of capitalisation errors which resulted from the settings used when macros or voice commands were created. Initially, these were considered to be software-specific errors and beyond the control of the respeakers. It was later realised that these errors were preventable and it is likely that when respeakers do avoid them at future events, the spread of edition and recognition errors will be very similar across the three corpora.

Where the RLE and RLE* corpora do differ from the data collected in Ofcom and LiRICS is in the seriousness of the errors. Figure 12 shows the total of

serious, standard and minor errors in each corpus:

	RLE	RLE* (excluding software errors)	Ofcom	LiRICS
Serious	2%	4%	5%	8%
Standard	12%	24%	39%	36%
Minor	85%	72%	56%	56%

Figure 12. Total serious, standard and minor errors

In all corpora, as the severity of the error decreases, the number of errors increases. However, in the live event data, the proportion of minor errors is far higher than that in either the Ofcom or LiRICS data, whilst fewer serious and standard errors are seen. Whilst this trend at live events is an advantageous one, since minor errors have less impact on the audience, further investigation is required to understand the reason behind it.

In Figure 13, the errors are classified by severity and type, i.e. edition or recognition:

	Edition				Recognition			
	RLE	RLE* (excluding software errors)	Ofcom	LiRICS	RLE	RLE* (excluding software errors)	Ofcom	LiRICS
Serious	5%	5%	2%	9%	1%	3%	7%	6%
Standard	38%	38%	42%	38%	2%	1%	31%	34%
Minor	57%	57%	56%	53%	97%	96%	62%	60%

Figure 13. Total serious, standard and minor errors categorised as edition or recognition errors

The general spread of edition errors was comparable across all corpora, although there was a higher proportion of serious edition errors at live events than in the Ofcom study. There is no clear reason for the raised number of serious errors, so each must be evaluated within the context in which it occurred.

The spread of recognition errors was very different in the live event corpus, where almost all errors were minor. One explanation is that collaboration between the respeakers and presenters meant that the respeakers were well-informed about the content of each event and the preparation time allocated enabled them to train in key vocabulary before the event began, thus avoiding many of the potential serious or standard recognition errors that might otherwise have occurred. Similarly, the speech rates experienced

at these events were lower than those in many programmes seen on television, which might have prevented more serious edition and recognition errors.

Another explanation lies in how the respeakers corrected errors during the events. When respeaking on television in the UK, respoken text is pulled and broadcast automatically, so most corrections, indicated with (--), must follow any error made; 17 such corrections were noted across the seven events. The software used at the live events allowed the respeakers more flexibility when making corrections. Streamtext, used at the two mobile events (four and seven) allowed respeakers to edit content that had already appeared on the audience's screen. The respeakers did not need to use the (--) on these occasions, as the corrected version replaced the original. Video footage of the events provides evidence of these corrections. In the Text on Top interface, used at the seated events, the respeakers are able to view and edit the titles before sending them to the audience's screen. It is likely that they did make some corrections in this interim interface, and these would not have been recorded on screen or in the transcript. The respeakers were all familiar with NER evaluation and the practice of trying to limit corrections to serious and standard errors, because of the latency caused by any correction made. If they were able to follow this practice at the live events, this could also be an explanation for why a higher proportion of minor errors appeared on screen, although we cannot state this with certainty. In future research, logging software would be a good way to examine this more closely.

Latency also impacts on quality and the viewing experience. Following the methods used in the Ofcom (2015) study, latency measurements were taken 2-3 times a minute, at intervals of 20-30 seconds. The time between a word being spoken and appearing on screen was recorded manually and when possible, phrase- and sentence-final words were used as these were often spoken more clearly. In the Ofcom study, the average delay was 7-8 seconds when respeaking without as-live cueing was used[9]. As shown in Figure 14, at the live events, the average latency in the samples ranged from 4.3 to 7.5 seconds, with 5.8 seconds being the average across all the events. This is lower than that seen in the Ofcom study, but still above the recommended 3 seconds. No data on latency was collected in the LiRICS pilot.

	Round One				Round Two		
Event	**1**	**2**	**3**	**4**	**6**	**7**	**8**
Latency in seconds:							
Average	6.7	7.5	5.9	4.5	5.8	4.3	5.7
Low	2.3	4.4	2.3	3.1	2.4	2.8	2.3
High	17.4	19.7	13.4	9.3	11.7	9.6	13.9

Figure 14. Latency at each event in seconds

The Ofcom study also recorded peaks of 10-21 seconds. Some of the peaks at the live events were also high, though still within the Ofcom range. What was noticeable at the live events is that the peaks were kept to a minimum. Sometimes the event itself facilitated this, by allowing the latency to be reset to zero, for example when video clips were played (event three) or the tour moved on to a new object (events four and seven). At some events, the presenters began to monitor the titles of their own accord, occasionally pausing to allow the titles to catch up.

It is noticeable that the lowest average latencies were seen at events where the Streamtext software was used (four and seven); here, titles are displayed directly onto the audience's screens, without the interim interface seen in Text on Top. The higher latencies seen with Text on Top may have been due to error correction or may have been the result of the respeakers having to remember to regularly cue out the text they had respoken. Whatever the cause, navigating this fine and important balance between accuracy and latency has a significant impact on title quality, as discussed further in section 4.7 below.

4.6 NER analysis: case studies

In order to better understand the transference of this service into the live event sector, a closer consideration of the data collected at individual events is needed. Two events from the second round of testing have been chosen for this purpose, since together they illustrate the complexities encountered at live events and the diversity of content and set-up seen. Event six is a post-screening Q&A discussion and event seven is a museum tour.

In addition to data on the spread of errors (Fig. 6 and Fig. 7 above), additional data will also be analysed, including speech, respeaking and titling rates, the reduction rate from the original content to the respoken content and the resulting change in lexical density. Figure 15, which contains two sets of data for events four and seven, reflects the changes that occurred as a result of drops in internet connectivity at the mobile events (see section 4.6.2):

	Round One					Round Two			
Event	**1**	**2**	**3**	**4**	**4 (Tablet view)**	**6**	**7**	**7 (Tablet view)**	**8**
Speech rate in clip WPM	143	135	136	123	123	145	112	112	133
Respeaking rate WPM	119	134	114	99	99	111	104	104	130
Title speed WPM	104	119	102	86	84	99	92	85	116
Reduction rate in titles	27%	12%	25%	30%	32%	32%	18%	24%	13%
Lexical density of spoken transcript	27%	37%	38%	34%	34%	25%	38%	38%	33%
Lexical density of respoken transcript	32%	39%	42%	41%	41%	29%	41%	42%	33%

Figure 15. Speed, reduction rate and lexical density at each event

4.6.1 Event six: Post-screening discussion

The accuracy of the post-film discussion of *The Piano* at the Depot in Lewes was the lowest of all the events: 97.3% (substandard), including software errors, and 98.4% (acceptable) excluding them. Given that this was one of the later events, a higher score might have been expected, yet it was at this event that the highest number of both edition errors and recognition errors occurred. What exactly does this score reveal about the access to the event and what was the impact on the audience's experience?

As with the other events, as the severity of the error decreased, the number of errors increased. There were 2 serious, 20 standard and 22 minor edition errors and 1 serious, 0 standard and 113 minor recognition errors, including software-specific errors. Excluding software-specific errors, the number of minor recognition errors fell to 41.

The high number of recognition errors is in part due to the macro settings used at this event; there were numerous capitalisation errors that could have been avoided (for example, the one seen in Fig. 16, segment 4, below). The high number of edition errors is reflected in the difference between the speech rate (145wpm) and the titling rate (99wpm) and captured in the reduction rate at the event, calculated as the percentage of words from the spoken transcript that are omitted in the respoken titles. The reduction rate at this event was one of the highest at 32%, a rate more than double that seen at the events with public speakers (two and eight), for example. Unlike those two events, event six was based on an audience discussion. Whilst there was a chair who opened the event with a series of questions, the audience were equally responsible for determining where the

conversation went. Consequently, the respeakers only had a very general notion of what content they would meet. As it happened, in the portion of the event analysed, the discussion moved from the storyline of *The Piano* to how music in films is made accessible.

Figure 16 contains segments of the NER analysis for event six, which illustrate a range of serious, standard and minor errors:

Segment	Spoken	Respoken	Comment
1	I can hear a piano better than hear a human voice. I know what's going to happen in a piano. I get it right. But I don't know what the next word is going to be from someone else so I've always got that difficulty.	I can hear a piano better than I can hear a human voice. I note I'm going to hear any piano. -- I know.	**1 Correction:** --know [I note]; no penalty applied. **1 Serious edition error (misinformation)**: 'I know I'm going to hear any piano is not necessarily' what 'I know what's going to happen in a piano' means. **2 Standard edition errors (content omitted):** But I don't know what the next word is going to be from someone else + so I've always got that difficulty.
2	I know she wasn't deaf… Um… I know she wasn't deaf, but she had a disability.	I know she was not deaf. I know she was not deaf but she had a disability.	**2 Correct editions:** Um…, repetition of 'I know she wasn't deaf'
3	She couldn't speak. Can't really work out the reason.	I can work out the reason.	**1 Standard edition error:** She couldn't speak. **1 Serious recognition error:** Can/can't. This is likely to be a recognition error, since the two words are similar. A new, plausible sentence is created.
4	It's, it's a kind of very small growing movement but it's all to do with the inclusion of, er, so media access professionals, so subtitlers and audio describers in the filmmaking process itself.	**JOSH**:, it is a small, growing movement, **B**ut it is to do with the **inclusion so subtitlersan** audio describers **I**n the music process themselves.	**5 Correct editions:** It's (doubled), a kind of, very, er, so **1 Serious edition error (misinformation):** 'music process' **3 Minor edition errors:** punctuation (,) (0.25);media access professionals (0.25); try to create a dialogue (0.25) **1 Minor recognition error:** it (Logged as a respeaker error, rather than software-specific error, as this was caused by a macro-setting) **5 Software-specific errors:** Highlighted in bold (all listed as minor recognition errors)
5	So it's not just, "Well, we've got subtitles there we go, um, we're, we're ticking boxes," but it's actually thinking about subtitling and media access in general as a new, as another cinematic tool in many ways.	It is about thinking about **subtitling and** media access in general **As** another cinematic tool.	**4 Correct editions:** But, actually, as a new, in many ways **1 Standard edition error:** omitting 'So it's not just, "Well, we've got subtitles there we go, um, we're, we're ticking boxes," **2 Software-specific errors:** Highlighted in bold and listed as minor recognition errors

Figure 16. Examples of errors at event six

In segment 1, the sentence 'I know what's going to happen in a piano' is respoken as 'I know I'm going to hear any piano'. On the first reading, there is a difference in meaning between the two sentences and, consequently, a penalty for a serious edition error was applied. However, in the video recording, it is notable that the person who has spoken these words pauses and reads what has been titled, before continuing, which seems to suggest that the titled content has been accepted as accurate since no attempt is made to correct it. Whilst the scoring was not adjusted, and this was left as a serious edition error, this illustrates the complexity involved in assessing respeaking at live events. Their interactive and participatory nature means that the speaker has the potential to affect error correction (and sometimes creation) in a way that is not possible in the recorded content broadcast on TV. When respeaking television, all that can happen is that the respeaker follows the content as it is broadcast. At live events, it is possible for respeakers and presenters to communicate directly and interact: respeakers can ask presenters to repeat content, for example by cueing [INAUDIBLE] or [PLEASE REPEAT THAT] and presenters have the opportunity to respond to the respoken titles they see.

The second serious error is a recognition error (segment 3), where 'can't' is respoken as 'can'. Since the resulting sentence, 'I can work out the reason' looks correct, the audience is misled. The final serious error occurred when 'film making process' was replaced with 'music process', most likely because the link between both industries was being discussed (segment 4). Whilst this error was not corrected, the audience had the opportunity to gain a fuller understanding of the filmmaking process as the discussion continued. So, although misleading, the impact of the error was not as serious as it might otherwise have been.

On the other end of the scale were the minor errors (segments 4 and 5), the most frequent kind in this event. The omission of 'media access professionals' whilst retaining 'subtitlers' and 'audio describers' is an example of a minor edition error since the audience received the basic information to understand what was being said, though an additional detail was lost. The misrecognition of 'and' as 'an' in segment 4 is an example of a minor recognition error as the audience are usually able to spot and understand the intended meaning.

Falling somewhere between these on the error scale, come the 20 standard errors that occurred in this event, all of which were edition errors, where the oral equivalent of a sentence is omitted. These errors often occur when the respeaker prioritises the central idea; for example, in segment 5, 'thinking about media access in general' is captured, but a comment about 'ticking boxes' by having subtitles is omitted. Sometimes a standard omission error might follow a correction, which seems to be the case in segment 1. In segment 3, a standard edition error can be observed, where 'She couldn't speak' precedes the serious edition error (can/can't) discussed

above. Although the serious error carries the greater penalty, in this instance, it seems that the standard error has a greater impact on the meaning, since it provides the connection between segments 2 and 3. For this reason, when possible, a respeaker should aim to correct both serious and standard errors.

As the reduction rate indicates, a significant percentage of the original was omitted. Yet, segments 1-5 and the analysis of the errors show that much of the key content of the event was respoken. How can both these statements be true? The high number of correct editions that were made at this event provides one explanation. As stated in section 4.5, correct editions occur when the respeaker omits or paraphrases the original, without content being lost. A total of 260 correct editions were made at event six, which was the highest number across all the events, the range being 87-260. The initial lexical analysis of the text supports the idea that much of the key content was retained[10]. The lexical density test is a readability test that tries to measure the proportion of words that have lexical content within a given text. The idea is that the higher the score, the harder the text as a whole is to read (UsingEnglish.com 2002-2019). At event six, the lexical density of the spoken transcript was found to be 25%, whilst that of the respoken transcript was higher, i.e. 29%. Although fewer words were respoken, and sentences were shorter, the lexical density actually increased. Taken along with the number of correct editions, this seems to indicate that the respeakers were able to preserve much of the complexity of the original discussion in a more compact form in the titles they created.

One of the broader aims of this research project was to help raise awareness about the importance of access and what respeaking involves. Since much of the discussion at this event focused on accessibility, it provided an excellent opportunity for this. Furthermore, the respeakers were in very close proximity to the audience, sitting, as they were, to the side of the screen, which meant the audience were very aware of their presence and the work they were doing. This seemed to be reflected in the feedback collected at the end of the event, where audience members were invited to complete a written questionnaire and comment on the event as a whole. When asked whether they preferred to attend events with subtitles, many audience members commented that although they did not need the subtitles themselves for access, they were happy to have them there for the benefit of those who did. Another noted that the respeakers must make on the spot decisions about what to include and stated, "the Q&A in the discussion were of high quality and increased my perception of respeaking."

Though subjective, as it comes from a single audience member, this comment suggests the respeaking was well received, something which was backed up by the responses to the question of whether the titles at the event were worse, as good as or better than the live subtitles audience members had seen on television. One person said it was worse, four said it

was as good and six said it was better. Two people chose not to respond, one saying they had not seen live subtitling on television to compare it to. No explanation was given for the 'worse' rating and the only comment about unsatisfactory access on the questionnaire referred to the limited scope of music labels often seen in pre-recorded subtitles.

Having positive audience feedback about the respeaking alongside the NER analysis is essential. Unlike television, where people are able to watch (sections of) programmes more than once, a live event cannot be rewound if something is missed. The audience want to engage fully and the respeaker must be able to navigate what is at hand to provide that access effectively.

4.6.2 Event seven: Museum tour

The seventh research event was a Highlights Tour at Manchester Art Gallery. The event was mobile and the tour stopped in numerous galleries. Audience members had individual tablets, on loan from Stagetext, through which they accessed the titles.

The accuracy at this event was high, with a score of 98.7% (good) including software errors and 99.0% (very good) excluding them. Unlike the post-screening discussion at event six, the core content of the museum tour was planned by the guide in advance and shared with the respeakers. They were able to enter and train specific items of vocabulary relating to the art work that would be viewed during the tour during their pre-event preparation. The tour was not scripted, and the audience were invited to ask questions, so there was also spontaneous speech to be respoken, but given the focus on particular works of art, the respeakers could prepare for this with a degree of prescience and very few vocabulary items posed problems during the tour.

A striking feature of the data for this event is the speech rate of 112wpm, and subsequent respeaking rate of 104wpm and title rate of 92wpm, which dropped to 85wpm, when the losses due to poor internet connection are included. Even without the losses, this is very low in comparison both with most of the other events seen in this study and TV content. The reason for this is the mobile nature of the tour. Whilst the tour itself lasted 50 minutes, time was needed to move around the gallery to the next item of interest. As with any audiovisual content, the audience must have enough time to process both the image and spoken word as 'the whole' comes from the combined content. When the spoken word is displayed visually in titles, more time is needed for this and when the focus of what is being said is the visual content, it is vital to ensure that time is left for the audience to see and explore that content after reading the titles, before new spoken content is introduced. While the words spoken may be uttered at regular speech rates, the pauses between blocks of speech reduce the average rate for the event.

It cannot necessarily be said, however, that these lower rates made it easier for the respeakers to provide access at this event. Given the mobile nature of the tour, the respeakers were not present alongside the guide and therefore had limited visual access to the content they were respeaking. Whilst they had been able to find online images of some pieces in advance of the tour, there were others that they had only read about. Nevertheless, they were able to respeak the full tour relying on the audio feed alone.

The fact that the audience had personal screens on which to view the titles also added to the complexity of the event. On the one hand, the number of tablets available limits the number of people for whom access can be provided; on the other hand, the tablets allow individuals to personalise the presentation of the titles, adjusting the size, colour, font and background to a combination of their choice. Time is needed before the event begins to inform the audience of this possibility and to allow them to explore the range of options available and become familiar with the settings. Logistically, this may mean extending the length of the event. Furthermore, in order to allow time for the audience to read the titles and have time to look at artwork, it may be necessary to schedule longer at each piece of work than in a non-accessible tour.

In this tour, the titles were streamed to the audience's tablets and this required a reliable internet connection at all locations in the tour and while moving between them. When reviewing the footage, it was noted that drops in internet connection had occurred which meant that not all the titles the respeakers were producing were being received by the audience as they toured the gallery. When a drop in internet connection occurred, no subtitles could be transmitted and any subtitles spoken during the drop were lost. When connectivity resumed, it was the most recent words that the respeaker had uttered that appeared on screen – the text created during the outage did not appear at all.

This impacted on the quality of access the audience received. Although the number of errors at this event remained low, the range of errors seen differed from that at other events, since serious and standard software-specific errors were recorded. At other events, the software-specific errors had been minor, resulting from unusual spacing and the occasional doubled letter. At this event, the software errors caused content to be omitted, which impacted on the coherence of the text produced. In places, what did appear could be mistaken as being poor editing by the respeakers, despite the fact that it was actually a technical fault. Take the example in Figure 17, where 'let me show you the sort of paintings the pre-Raphaelites' is omitted as a result of the drop in connectivity:

Segment	Spoken	Respoken content (excluding software-specific errors)	Audience tablet feed (including software-specific errors)
1	I showed you the pictures of Joshua Reynolds, just let me show you the sort of paintings that the pre-Raphaelites admired from the past.	I showed you the pictures of Joshua Reynolds, let me show you the sort of paintings the pre-Raphaelites admired from the past.	I showed you the pictures of Joshua Reynolds,admired from the past.

Figure 17. Example of software-specific errors seen at mobile events

This omission creates two recognition errors. Firstly, there is a missing space between 'Reynolds,' and 'admired'; more significantly, a sentence with new meaning is created, since now the audience are led to understand they will be looking at pictures Joshua Reynolds admired. In the actual respoken content, there were 27 edition errors, of which 3 were serious, 11 were standard and 13 were minor, and only 15 recognition errors, all of which were minor. In the content streamed to the tablet recorded during the tour, 26 edition errors were found, of which 3 were serious, 11 were standard and 12 were minor, and 23 recognition errors, including 1 serious error, 6 standard errors and 16 minor errors. The reduction rate also increased as a result of the dropped titles, rising from 18% to 24% on the tablets.

Since the content that was respoken during the drops never made it to screen, the effect of these drops was not incorporated in the latency calculation. It was instead noted in the event-specific comment made within the NER analysis and used as an opportunity for learning and improvement within the action research process.

In a mobile tour, it is therefore essential to check the internet connection throughout the building in advance in order to ensure that the audience will be able to receive the full content of the respoken titles. In certain buildings, this may either restrict the areas the tour can move through, or mean that a mobile boost is required to ensure the connection remains stable.

The fact that the respeakers are working remotely also has implications for how the guide (presenter) must conduct the tour. The audio will need to be shared with the respeakers via a mobile phone, either handheld, or through a headset. Allowing the presenter to get used to this before the tour begins is of great benefit. At any event, it is helpful to have a nominated person in charge of monitoring the titles and communicating any issues with the presenters. The presenters also need to understand how they can facilitate the work of the respeaker. In a museum tour, this is all the more important as the guide is unlikely to have their own tablet. Communicating with the audience via tablets is a different experience for any presenter and, in a tour, audience and presenter are in particularly close proximity. Repeating

questions from the audience and leaving time for responses and viewing time also mean the flow of the tour is a different one.

In practice, guidance like this is important for presenters at all events made accessible through respeaking. In advance of each event, the presenters received written information about what respeaking involved and how they could prepare to be respoken. This written guidance was updated between the two rounds of research events, and the suggestions were tailored specifically to different event types (presentations vs mobile tours, for example). There was also an opportunity for the presenters and respeakers to meet before the event began and to test out the respeaking. However, by the end of the research events, it had become clear that this was not enough. Presenters, and ideally all those people closely involved in organising respoken events, needed direct experience of both being respoken and accessing spoken content through respeaking and the opportunity to reflect on the experience before the day of the event, so that this could feed into their preparation for the actual event. Training for presenters was developed for this purpose, which is detailed further below.

4.7 Discussion

The NER analysis of the research events reveals that industry-standard accuracy can be achieved through respeaking at live events. However, the original question also sought to explore what training and workplace procedures need to be in place to ensure quality.

The case studies of events six and seven reveal how important the interaction between all those involved at the event is, not simply from a logistical point of view in organising the event, but in the effective running of the event. In order to work in the live event setting, respeakers will have to expand their repertoire and take on new skills; they will need knowledge of the equipment they are using (see section 4.2) and will need to be able to set it up and troubleshoot independently. They will also need to act to a certain extent as an "accessibility expert" (Remael *et al.* 2019), educating and informing on-site staff of the role they must play to facilitate the audience's experience of the event as a whole. This will be incorporated into the future and finalised Respeaking at Live Events training programme.

Through the live event testing, it also became clear that the guidelines for presenters and venues were not sufficient. In the same way as respeakers pass a confidence threshold in their work, so, too, must the people who are involved in facilitating the event. Deaf awareness training would be useful for all staff; 'respeaking awareness training' should be provided for all presenters and tailored to the person and type of event. The experience of a tour guide will be very different from that of a conference presenter, yet both need to understand the importance of allowing the audience time to respond to a question that has been asked or to view visual content. For the guide, this may mean remembering to pause more often; for the

conference presenter, it might involve adjusting the format of his/her slides so that the titles do not block key content.

TV guidelines recommend that a subtitle should be moved so that the visual image is not obscured (Ford Williams 2009). An equivalent adage at live events would be that the respeaker should ensure the audience has time to watch key content. Whilst certain live event conventions were established during the initial training, these focused on the audio content. For example, when the respeaker could not make out what the presenter had said, they cued (INAUDIBLE), as a request for the presenter to repeat the missing content, allowing it to be captured. What if there is visual content the audience may miss? In the eighth event, the presenter spoke at a fairly rapid pace and included many gestures as he spoke. Should the respeaker have stepped in and sent a cue to the audience to (LOOK UP)? Whilst this would not be practical on live TV, in the live event setting, the respeaker could take this step and enter even more fully into the dynamism of the event. This is certainly something for further consideration and discussion at particular events. Similarly, where an event relies on videos or images, the respeaker may need to work with the presenter in advance to develop a smooth handover routine. On the news, it is common to see the image of the new story with the subtitle from the previous one still scrolling across screen. Respeakers at events may be able to avoid this with some communication with the presenter.

The latency at live events was also discussed briefly in section 4.5 above. Whilst falling within television ranges, the ideal scenario would be to reduce latency to an absolute minimum. Allowing occasional moments for the audience to pause and reflect, while the titles catch up, could be invaluable. Whilst the option exists on television for a broadcast delay, whereby the visual is delayed so that it is sync with the audio, this is not an option at a live event; communication with the presenter and finding a natural rhythm for pauses may be a solution to this.

5. Conclusions

Although there are many misconceptions about respeaking and criticisms of the errors it can produce, this study has shown that experienced respeakers were able to move into the live event setting and produce good quality access, in line with the industry-standard NER model, across two series of diverse events.

The Respeaking at Live Events training programme was developed to map out this process. By embedding it within a framework of action research, it was possible to involve professionals alongside academics and audience members and to proceed by continually evaluating and building on what had gone before. Throughout the project, there were opportunities to test ideas, reflect on the procedures used and access provided, make improvements, and repeat the testing to see what progress had been made.

Ultimately, this meant that examples of best practice for live events were found that could be applied to diverse scenarios. Furthermore, since everyone is participating in real-time during an event, there are always opportunities for contact and discussion. Maximising these felt like a very natural process and an effective methodology as everyone involved was able to add their expertise to the research.

Respeaking is skilled work, though this is something that is not always recognised. Developing certifiable pathways towards professionalisation and continuous professional development will be one way to change this (Romero-Fresco *et al.* forthcoming). LiRICS has developed a certification process for respeakers on television and at live events with a view to consolidating the certified status of this profession. It is hoped that the findings presented here could contribute to this. As discussed in section 3, there is a clear overlap in audience expectations for live (sub)titling in both settings, so any experience gained of working in live events will only serve to complement the portfolio of a television respeaker.

Another aim of this research was to increase audience awareness, and hopefully appreciation, of respeaking. Although not without criticism, the audience who attended these research events were impressed with the quality of access the service offered, especially as the events progressed and procedures were tightened. We live in a world where speech recognition is being used ever more widely, from Siri and Cortana on our phones and computers, to Alexa in our homes. Respeaking is suddenly becoming more understandable and perhaps now is the time for its use to be re-evaluated.

Certainly, the research events provided an opportunity for the audience to see respeaking in action, and to be involved in its success, in a way that cannot happen on television. The 'liveness' of events is in the presence of those involved and the interaction they bring: the audience, along with the respeakers, presenters and venues, become direct participants in what is happening; every person, whatever their hearing status and whatever their role on the day, needs to understand why access matters, why respeaking in particular matters and what is involved in providing high quality live titles. This will help to manage the audience's expectations of respeaking and lead to better access provision.

In this way, access is not just something that benefits everyone involved, but it also depends on everyone involved being actively engaged in it. Undoubtedly, this live context necessitates increased dialogue, communication, interaction and understanding between all stakeholders, which may prove challenging at times. However, it could also lead to respeakers having the opportunity to tailor the access they are providing to the specific situation and to the audience present. This is equity in action.

If the Universalist account of accessibility states that access concerns all human beings (Greco 2018) and holds the user at its heart, the approach

suggested in this paper goes a step further by providing a model and an invitation for everyone to be actively engaged and to participate in the access as it is provided.

Acknowledgements

I would like to thank everyone involved in the running of these events – the venues, respeakers, presenters, interpreters and people helping on site, without whom the events could not have taken place, the audience members, who volunteered their valuable time, and Stagetext, for their support throughout this research project and for the use of their tablets at the museum tours.

In particular, I would like to thank Pablo Romero-Fresco and Lucile Desblache for their support, advice and guidance throughout this project and Kate Dangerfield, Gian Maria Greco, Hayley Dawson and Dawn Jones for the thoughtful discussions we have had on these topics.

References

- **CCAC** (2016a). *What is CART and FAQs – Find a CART Captioning Provider.* http://ccacaptioning.org/faqs-cart (consulted 8.4.2019).

- **CCAC** (2016b). *Support for Live Event Captioning: Apply for CCAC Grant.* http://ccacaptioning.org/ccac-sponsorship-for-cart-a-new-captioning-advocacy-program (consulted 8.4.2019).

- **CRTC** (2019). *Broadcasting Notice of Consultation CRTC 2019-9*. Canadian Radio-television and Telecommunications Commission. https://crtc.gc.ca/eng/archive/2019/2019-9.htm (consulted 10.11.2019).

- **Cock, Matthew, Molly Bretton, Anna Fineman, Richard France, Claire Madge and Melanie Sharpe** (2018). *State of Museum Access 2018 – Does Your Museum Website Welcome and Inform Disabled Visitors?* http://www.stagetext.org/uploads/State_of_Museum_Access_2018.pdf (consulted 8.4.2019).

- **Díaz Cintas, Jorge and Aline Remael** (2007). *Audiovisual Translation: Subtitling.* Manchester: St Jerome.

- **EUR-Lex** (2019). *Directive (EU) 2019/882 of the European Parliament and of the Council of 17 April 2019 on the Accessibility Requirements for Products and Services (Text with EEA relevance)*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2019.151.01.0070.01.ENG&toc=OJ:L:2019:151:TOC (consulted 9.11.2019).

- **Eichmeyer, Daniela** (2017). "Speech-to-text interpreting: Barrier-free access to universities for the hearing impaired." *Proceedings of the 1st Swiss Conference on Barrier-free Communication.* Winterthur: ZHAW, 6-15.

- **Eugeni, Carlo** (2006). "Introduzione al rispeakeraggio televisivo." *InTRAlinea Special Issue: Respeaking.* http://www.intralinea.org/specials/article/Introduzione_al_rispeakeraggio_televisivo (consulted 7.4.2019).

- **Ford Williams, Gareth** (2009). *Online Subtitling Editorial Guidelines* V1.1. BBC.

http://www.bbc.co.uk/guidelines/futuremedia/accessibility/subtitling_guides/online_sub_editorial_guidelines_vs1_1.pdf (consulted 23.9.19).

- **Greco, Gian Maria** (2018). "The nature of accessibility studies." *Journal of Audiovisual Translation* 1(1), 205-232.

- **Lambourne, Andrew** (2006). "Subtitle respeaking a new skill for a new age." *InTRAlinea Special Issue: Respeaking.* http://www.intralinea.org/specials/article/Subtitle_respeaking (consulted 19.2.2018).

- **Legislation.gov.uk** (2010). *Equality Act 2010,* 15,2. http://www.legislation.gov.uk/ukpga/2010/15/contents (consulted 7.4.2019).

- **Mann, Blair** (2014). *Equity and Equality Are not Equal.* https://edtrust.org/the-equity-line/equity-and-equality-are-not-equal (consulted 10.4.2019).

- **Mele Scorcia, Antonia** (2018). "Surtitling and the audience: A love-hate relationship." *The Journal of Specialised Translation* 30*,* 181-202.

- **McIntyre, Dan, Zoe Moores and Hazel Price** (2018). *Respeaking Parliament: Using Insights from Linguistics to Improve the Speed and Quality of Live Parliamentary Subtitles.* Huddersfield: Language Unlocked/Institute for Applied Linguistics.

- **Moores, Zoe and Pablo Romero-Fresco** (2015). "The language of respeaking." Paper presented at *5th International Symposium on Respeaking, Live Subtitling and Accessibility,* (UNINT Rome, 12 June).

- **Nertrial.com**. n.d. http://nertrial.com/Scoresheets.htm (consulted 28.5.19).

- **Ofcom** (2013a). *Measuring the Quality of Live Subtitling*. https://www.ofcom.org.uk/__data/assets/pdf_file/0017/51731/qos-statement.pdf (consulted 7.4.2019).

- **Ofcom** (2013b). *The Quality of Live Subtitling: A Consultation.* https://www.ofcom.org.uk/consultations-and-statements/category-1/subtitling (consulted 7.4.2019).

- **Ofcom** (2014). *Measuring Live Subtitle Quality: Results from the First Sampling Exercise.* https://www.ofcom.org.uk/__data/assets/pdf_file/0019/45136/sampling-report.pdf (consulted 10.11.19).

- **Ofcom** (2015). *Measuring Live Subtitle Quality: Results from the Fourth Sampling Exercise.* https://www.ofcom.org.uk/research-and-data/tv-radio-and-on-demand/tv-research/live-subtitling/sampling_results_4 (consulted 11.9.2017).

- **Ofcom** (2017). *Ofcom's Code on Television Access Services.* https://www.ofcom.org.uk/__data/assets/pdf_file/0020/97040/Access-service-code-Jan-2017.pdf (consulted 19.2.2018)

- **Reason, Peter and Hilary Bradbury** (2008). *The SAGE Handbook of Action Research Participative Inquiry and Practice.* 2nd Ed*.* London: SAGE Publications.

- **Remael, Aline, Pilar Orero, Sharon Black and Anna Jankowska** (2019). "From translators to accessibility managers: How did we get there and how do we train them?" *MonTI* 11, 131-154.

- **Romero-Fresco, Pablo** (2011). *Subtitling Through Speech Recognition: Respeaking.* Manchester: St Jerome.

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

- **Romero-Fresco, Pablo** (2018a). "In support of a wide notion of media accessibility: Access to content and access to creation." *Journal of Audiovisual Translation* 1(1), 187-204.

- **Romero-Fresco, Pablo** (2018b). "Respeaking: Subtitling through speech recognition." Luis Pérez-González (ed.) *The Routledge Handbook of Audiovisual Translation Studies.* London: Routledge, 96-113.

- **Romero-Fresco, Pablo** (*forthcoming*). "Negotiating quality assessment in media accessibility: The case of live subtitling". *Universal Access in the Information Society.*

- **Romero-Fresco, Pablo and Juan Martínez** (2015). "Accuracy rate in live subtitling: The NER model." Jorge Díaz-Cintas and Rocío Baños-Piñero (eds). *Audiovisual Translation in a Global Context: Mapping an Ever-changing Landscape.* London: Palgrave MacMillan, 28-50.

- **Romero-Fresco, Pablo, Sabela Melchor-Couto, Hayley Dawson, Zoe Moores and Inma Pedregosa** (*forthcoming*). "Respeaking certification: Bringing together training, research and practice." *Linguistica Antverpiensia New Series.*

- **Szarkowska, Agnieszka, Krzysztof Krejtz, Łukasz Dutka and Olga Pilipczuk** (2018). "Are interpreters better respeakers?" *Interpreter and Translator Trainer* 12(2), 207-226.

- **Stagetext** (2011). *A Play in the Life of a Captioner.* https://www.youtube.com/watch?v=ZIfNzSt5rSw (consulted 7.4.2019)

- **Udo, John-Patrick and Deborah I. Fels** (2010). "The rogue poster children of universal design: Closed captioning and audio description." *Journal of Engineering Design* 21(2-3), 207-221.

- **UsingEnglish.com** (2002-2019). *English Grammar: Lexical Density Test.* https://www.usingenglish.com/glossary/lexical-density-test.html (consulted 11.11.19)

- **Vanderplank, Robert** (2016). *Captioned Media in Foreign Language Learning and Teaching: Subtitles for the Deaf and Hard-of-Hearing as Tools for Language Learning*. Basingstoke: Palgrave Macmillan.

Biography

Zoe Moores is an AHRC TECHNE-funded PhD research student and Visiting Lecturer at the University of Roehampton. Her research explores how respeaking could be introduced into the live event setting in the UK to broaden the access provided for d/Deaf, deafened and hard of hearing audience members and the wider audience. She is part of the GALMA research group and is involved in a number of accessibility-related projects. Zoe worked as a respeaker and subtitler at Ericsson and continues to translate and subtitle on a freelance basis.

Email: z.moores@roehampton.ac.uk

Notes

[1] One current project is the Interlingual Live Subtitling for Access project (ilsaproject.eu).
[2] LiRICS is an accreditation scheme, currently being ratified, which will provide recognised certification, and therefore professional status, for respeakers working in English across television and live events. More information can be found in Romero-Fresco (forthcoming) and at: http://galmaobservatory.eu/services/certification
[3] At events where audience members attend in person, there is also the possibility that additional audience members attend remotely, for example, through conference call software. Given the broad scope of live events defined above, the primary focus of this project has remained on non-remote events.
[4] *Supratitles* or *surtitles* are the titles that appear on screens that are suspended above or at the side of the stage during theatre and musical performances. In the UK, Stagetext refers to these same titles as *captions*.
[5] Stagetext refers to these same titles as *live subtitles* to differentiate them from their pre-prepared captions for theatre.
[6] These figures, provided by Stagetext, were drawn from their yearly Trustee's Reports and Accounts documents.
[7] The research for this project was submitted for ethics consideration under the reference MCL 15/ 025 in the Department of Media, Culture & Language and was approved under the procedures of the University of Roehampton's Ethics Committee on 27.1.16.
[8] The film on screen is *Blue Pen* by Vital Xposure, directed by Julie McNamara (pictured) and Caglar Kimyoncu.
[9] As stated in section 2.1, in the Ofcom study, lower average latencies were seen when parts of the programme were cued out as-live. Since the events were respoken and not cued, the figure for pure respeaking is used here to allow a clearer comparison to be made.
[10] Lexical density was calculated using the text analyser on the usingenglish.com website (https://www.usingenglish.com/resources/text-statistics.php).