



# Measuring live subtitling quality

Results from the fourth sampling exercise

Publication date: 27 November 2015



# Contents

| Section      |                                 | Page        |
|--------------|---------------------------------|-------------|
| 1            | Summary                         | 2           |
| 2            | Accuracy                        | 7           |
| 3            | Latency                         | 15          |
| 4            | Speed of subtitling             | 19          |
| 5            | Edition rates                   | 26          |
| 6            | Next steps                      | 29          |
| <b>Annex</b> |                                 | <b>Page</b> |
| 1            | External review of measurements | 32          |

## About this document

This document is the last of four reports on the quality of live subtitling of British television programmes, based on samples drawn from live-subtitled programming broadcast in April and May 2015 by the BBC, ITV, Channel 4, Channel 5 and Sky. In addition to setting out the results of measurements of accuracy, latency and speed for the samples, the report sets out the next steps that Ofcom intends to take. This includes research on how subtitle users would prefer trade-offs to be addressed.

## Section 1

# Summary

## Introduction

1.1 This document is the last of four reports on the quality of live subtitling on British television programmes, based on samples drawn from live-subtitled programming broadcast in April and May<sup>1</sup> 2015 by the BBC, ITV, Channel 4, Channel 5 and Sky.

## Background

1.2 Ofcom's statutory duties require it, among other things, to provide guidance to broadcasters on how they should promote the understanding and enjoyment of their services by people with sensory impairments, including people with hearing impairments.

1.3 In the light of our statutory duties and of continuing concerns about the quality of live subtitling, in 2013<sup>2</sup> Ofcom began a two-year project to measure the quality of live subtitling, in order to identify areas for improvement and encourage broadcasters to act upon these.

1.4 For this purpose, Ofcom asked five broadcasters<sup>3</sup> to measure the following dimensions of quality on samples of their live-subtitled programmes selected by Ofcom every six months:

a) accuracy: the number and type of errors (i.e. minor spelling errors, major omissions or factually misleading subtitles), using the NER model<sup>4</sup>;

b) the average latency of the subtitling (the delay between speech and live subtitling), and the range of latencies; and

c) the average speed of the subtitling, measured in words-per-minute ('wpm') rates.

1.5 The first three reports were published between April 2014 and May 2015<sup>5</sup>. For the purpose of this project, we asked a team of experts from the University of Roehampton to check the broadcasters' measurements for consistency.

## Fourth report

1.6 Sections 2, 3 and 4 of this document report on the three dimensions of live subtitling quality listed above, as well as on issues raised by the reviewers, such as the number of peaks in subtitling speed. Section 5 reports on the extent to which subtitles are edited down.

---

<sup>1</sup> With the exception of two samples of entertainment programmes shown on Channel 5, which were broadcast in June of the same year.

<sup>2</sup> *Measuring the quality of live subtitling*, Ofcom, 16 October 2013 (<http://stakeholders.ofcom.org.uk/consultations/subtitling/statement>)

<sup>3</sup> The broadcasters involved in this project were: the BBC, ITV, Channel 4, Channel 5, and Sky.

<sup>4</sup> For more about the NER Model, see: Ofcom, *Measuring live subtitling quality – Results from the first sampling exercise*, 30 April 2014; paragraphs 2.2-2.5.

<sup>5</sup> All four reports on the quality of live subtitling can be found at the following link: <http://stakeholders.ofcom.org.uk/consultations/subtitling/>

- 1.7 Section 6 sets out the next steps that will follow this report.
- 1.8 The qualitative report by the external reviewers from the University of Roehampton, including a survey of the main findings from the four sampling exercises, can be found at Annex 1.
- 1.9 In particular, the external reviewers noted that, in the fourth round of sampling:
- a) the average **accuracy** rate was the highest of all four sampling rounds, at 98.55%. This was mostly due to the continued use of a mix of live subtitles and subtitles prepared in advance for scripted or pre-recorded segments; but it also reflected fewer technical faults. This round of measurements also saw the highest proportion (80%) of samples deemed to have subtitles of acceptable quality, with accuracy rates at or above 98%. However, these accuracy rates were not adjusted to take account of the impact of rapid subtitling (see paragraphs 1.16 to 1.19 below);
  - b) the average **latency** (5.6 seconds) was broadly in line with the first two reports (5.4 and 5.6 seconds respectively), but higher than the third report (5.1 seconds), and well above the 3 seconds recommended in Ofcom's Guidelines on the provision of television access services<sup>6</sup>. This is despite a reduction in the average latency in samples of news bulletins with respect to the first two reports; and
  - c) the average subtitling **speed** was below the range of 160-180 wpm recommended for pre-recorded subtitles in all but one sample and so acceptable. However, a very high proportion (92%) of samples featured short bursts of 'rapid subtitles' (i.e. instances of live subtitling at speeds exceeding 200 wpm), although smaller than in the third round (99%). This issue affected news bulletins more than other genres: 87% of the samples of news included at least one example of subtitles faster than 200 wpm, and in 20% of the samples at least half the subtitles were faster than this.

## Tackling latency

- 1.10 Since our 2013 consultation<sup>7</sup>, we have had extensive discussions with broadcasters about whether the apparently intractable problems of latency could be addressed by inserting short delays in the transmission of certain programmes that are subtitled live. Broadcasters argued that the cost and complexity of inserting short delays would be disproportionate to the benefits that subtitle users would enjoy.
- 1.11 However, the BBC and Channel 4, working with their access service and playout provider, Ericsson, have identified an alternative technical solution that could reduce latency significantly – so-called 'switchable delays'. In brief, this would take advantage of delays that are inherent in the coding, multiplexing and playout processes for video, audio and subtitling. By allowing subtitles (but not audio or video) to short circuit these processes, it may be possible to reduce latency by 4-5 seconds in channels broadcast in high definition (HD).
- 1.12 Ofcom understands that the trials conducted in the BBC R&D lab have yielded some promising results. Unlike the proposal for short delays, the benefits would extend to all live subtitled programmes, including news and sports. More detail on this,

<sup>6</sup> see *Ofcom's Code on Television Access Services*, 13 May 2015 (<http://stakeholders.ofcom.org.uk/binaries/broadcast/other-codes/tv-access-services-2015.pdf>); Annex 4.

<sup>7</sup> Ofcom, *The quality of live subtitling – Improving the viewer experience*, 17 May 2013 (<http://stakeholders.ofcom.org.uk/binaries/consultations/subtitling/summary/subtitling.pdf>)

including the BBC's plans to roll it out on their HD channels, is given in Section 3. Initially, this technical solution will only be available on HD channels, though there is the prospect of making it available on SD channels.

- 1.13 Channel 4 has told us that it is planning a proof-of-concept trial of a very similar approach. The BBC and Channel 4 have told us that they are willing to share the results of their trials, and we will seek to facilitate discussions with ITV, Channel 5 and Sky on the technical approach, so that they consider whether it would be feasible in their particular circumstances. But we recognise that broadcasters use different technical systems and that this approach may not work for all.
- 1.14 The reduction in latency resulting from these plans would not tackle the particular problem of lower subtitling accuracy scores for some popular entertainment and chat shows. Some of these, like Gogglebox, are pre-recorded, and Ofcom believes that broadcasters should make greater efforts to enable subtitles for such programmes to be prepared in advance, even if they have to be cued out live.
- 1.15 At a recent meeting of the Eurovision Access Services Experts Group<sup>8</sup>, the Flemish public service broadcaster (Vlaamse Radio- en Televisieomroeporganisatie, or VRT) demonstrated to Ofcom and others how it uses short delays to produce higher quality subtitles. VRT now produces about 20 hours a week of programmes with 'delayed live subtitling'. While we remain of the view that short delays in live transmissions would be technically feasible in the UK, we consider that it is likely to be more fruitful in the medium term for UK broadcasters to concentrate their efforts on using switchable delays to secure significant improvements in latency, where their transmission infrastructure allows.

### **Rapid subtitling**

- 1.16 As with the third report, this report includes 'adjusted' accuracy measurements which treat subtitles shown at speeds above 200 wpm as standard errors for the purpose of accuracy measurements. This reflects research (see Section 4) which suggests that subtitles shown at speeds above 200 wpm are too fast for subtitle users.
- 1.17 The results are summarised in Section 2. In short, once the median scores are adjusted on that basis, the overall average accuracy falls slightly below the acceptability threshold, at 97.93%. At the same time, the proportion of samples with accuracy below 98% increases from 20% to 50%.
- 1.18 In our third report, we said that we would ask broadcasters to report on what measures they were taking or planning to avoid rapid subtitling. As explained in Section 4, all broadcasters have told us that they impose software-enabled limits on the speed at which subtitles are shown on screen, but that none of them uses a limit as low as 200 wpm out of a concern that this could result in higher latency.
- 1.19 Some broadcasters indicated that they are considering other approaches alongside speed limits, such as specific training for subtitlers, or the improvement of the broadcast infrastructure to lessen the impacts of 'bottlenecks' for the live subtitle data stream.

---

<sup>8</sup> see: <https://www3.ebu.ch/contents/events/2015/10/eurovision-access-services-exper.html>

## Next steps

- 1.20 In our 2013 Statement, we said that at the end of this project we would consider whether a review of the Guidelines was necessary, in the light of the evidence collected through the sampling exercises. Given the developments in subtitling approaches and understanding of quality issues since the Guidelines were first published, we consider that a review would be worthwhile, and we will also consider whether changes should be made to rules in the Code on Television Access Services.
- 1.21 However, before proposing any specific changes, we intend to discuss the main findings of the four rounds of sampling, as well as any additional research carried out in recent years, with groups representing subtitle users and with broadcasters and access service providers. For this reason, any conclusions that could be drawn at this stage are preliminary.
- 1.22 That said, it is clear from this exercise that:
- a) by making every effort to prepare subtitles in advance for pre-recorded programmes, broadcasters and access service providers can significantly improve the quality of the viewing experience for subtitle users. We hope that programme producers will recognise the value of giving subtitlers sufficient time to prepare subtitles in advance, as live subtitling impairs the quality of their product for subtitle users;
  - b) broadcasters have helped to improve the quality of live subtitling for news programmes by giving access service providers access to relevant material, including running orders, autocue scripts and live feeds from the newsroom;
  - c) when producing live subtitling, trade-offs need to be struck between different quality dimensions. For example, it would be possible to produce accurate, verbatim subtitles for a fast-paced programme, but only if viewers accepted significant delays in the subtitles appearing; and
  - d) technical problems occur along the subtitling supply chain relatively frequently, although they affect a very small proportion of output. Not all technical problems stem from consumer equipment. Ofcom will consider how best to encourage broadcasters to limit the number of technical problems to a low level.
- 1.23 How trade-offs between different dimensions of quality are struck can have a significant impact on the viewing experience for subtitle users. We understand that broadcasters and access service providers are aware of these trade-offs, and have discussed some of them with subtitle users and groups representing their interests.
- 1.24 Nevertheless, we consider that it would be worthwhile commissioning independent research on subtitle users' attitudes towards these trade-offs, as:
- a) it would shed light on the preferences of subtitle users with different levels of hearing loss, different ages and different educational attainments, so that we can see whether there any differences in how they view the trade-offs; and
  - b) the outcome will be published on Ofcom's website, so will be transparent to all interested parties.



- 1.25 We have invited broadcasters, access service providers and groups representing the interests of subtitle users to comment on the issues that the research should look at, so that we can take their views into account before we commission the work.
- 1.26 We expect to publish the outcome of the research in early 2016, and to discuss it with stakeholders thereafter. We will take account of their views, the independent research we will commission, as well as recent third party research, before consulting on any changes to Ofcom's Guidance and Code.

## **Acknowledgments**

- 1.27 Ofcom would like to thank both broadcasters and subtitling providers for their collaboration in this project, as well as the external reviewers from the University of Roehampton.

## Section 2

# Accuracy

- 2.1 Live subtitling – especially when done through respeaking – is inevitably marred by mistakes, due to the complexity of the respeakers' task and to the limitations of recognition software<sup>9</sup>. These errors force viewers to spend time working out what was meant. In some situations, even tiny mistakes, such as the substitution of 'fifteen' for 'fifty', may cause confusion for the viewer and generate false but credible information.
- 2.2 Subtitling accuracy refers to the ability of the subtitles to correctly and fully convey the audible information that would not otherwise be accessible to subtitle users. A well-recognised method for the measurement of subtitling accuracy, which has been adopted by Ofcom for the purposes of this project, is the NER Model<sup>10</sup>. Throughout this project, the team of external reviewers identified 98% as the quality threshold above which the quality of subtitling can be considered as 'acceptable'.
- 2.3 As part of the research into subtitle users' attitudes towards live subtitling that we are planning to commission, we shall try to find out more about what accuracy levels, as measured by the NER Model, are considered acceptable by subtitle users.

## Summary of results

- 2.4 The figures in this Section illustrate the results of measuring accuracy on both an 'unadjusted' and 'adjusted' basis. The unadjusted results show the accuracy of all subtitles on the same basis, regardless of whether they are displayed at or below 200 wpm.
- 2.5 Research suggests that subtitles above 200 wpm ('rapid subtitles') will be unreadable for many subtitle users<sup>11</sup> (although we note the BBC's latest research in this area – see section 4). In order to account for the detriment in the readability of the subtitles resulting from rapid subtitles, since the third round of sampling we have included 'adjusted accuracy measurements' in our reports. The adjusted measurements treat rapid subtitles as 'standard' errors, which results in the accuracy score being reduced. For the purpose of the application of the NER model, standard errors are defined as those errors that "result in the omission of an information unit from the original text"<sup>12</sup>. Despite these errors do not actively create new misleading meaning, standard errors disrupt the flow/meaning of the original text and often cause surprise to the viewer<sup>13</sup>.
- 2.6 In both cases, 98% acceptable quality threshold is shown with a bold blue line.

<sup>9</sup> For more information on how subtitles are produced, see: Ofcom, *The quality of live subtitling – improving the viewer experience*, 15 May 2013, section 4.

<sup>10</sup> For more about the NER Model, see: Ofcom, *Measuring live subtitling quality – Results from the first sampling exercise*, 30 April 2014; paragraphs 2.2-2.5.

<sup>11</sup> Romero-Fresco, Pablo: *Quality in Live Subtitling: The Reception of Respoken Subtitles in the UK* (<http://connection.ebscohost.com/c/articles/78120106/quality-live-subtitling-reception-respoken-subtitles-uk>).

<sup>12</sup> Romero-Fresco, Pablo & Martinez, Juan: *Accuracy Rate in Live Subtitling – the NER Model* (2011), Roehampton University, UK

(<http://roehampton.openrepository.com/roehampton/bitstream/10142/141892/1/NER-English.pdf>)

<sup>13</sup> *ibid.*

## Unadjusted results

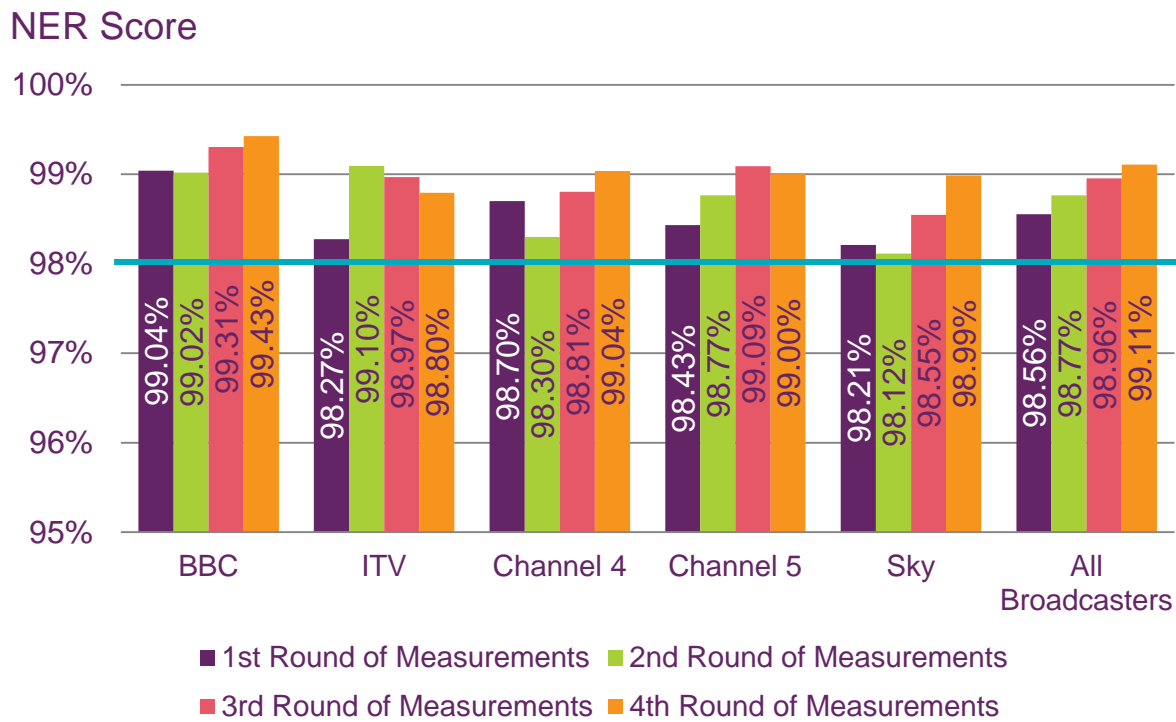
- 2.7 Figures 1, 2 and 3 below show the median unadjusted accuracy rates for all three sets of samples taken from news bulletins, chat shows and entertainment programmes in all four round of sampling.
- 2.8 This sampling exercise saw the highest proportion of samples with subtitling deemed as 'acceptable' or better since the beginning of this project (80%)<sup>14</sup>, and the highest proportion of samples with subtitling deemed as 'very good' or 'excellent' (29%)<sup>15</sup>. Furthermore, in the fourth round of measurements, two genres – news and chat shows – experienced their highest ever median accuracy rate, with samples of chat shows obtaining a median rate above 98% for the second time since the start of this project.
- 2.9 This round of measurements also recorded the highest median accuracy scores in samples chat shows and news bulletins. In the case of chat shows, the median accuracy (98.38%) was above the acceptability threshold for the first time since the first sampling exercise. As explained by the external reviewers, this was helped by a lower average speech rate (185 wpm) recorded in samples of chat shows than in the previous round of measurements (222 wpm).
- 2.10 The median accuracy in samples from entertainment shows was the lowest since the beginning of the exercise (98.44%). Almost one in three samples (30%) had an accuracy rate below the acceptability threshold, more than in the past two sampling rounds (respectively, 11% and 25%).
- 2.11 We note that, as the external reviewers pointed out, this is partially influenced by the different types of entertainment programming sampled over the course of this project. The reviewers note that some of the programmes selected in the later rounds of sampling, such as Game Changers (Sky) or Big Brother (Channel 5), share many traits with chat shows, such as higher average speech rates, or several people speaking at the same time.
- 2.12 One of the lessons learned from this exercise (see Section 6) is that providing subtitlers with scripts and pre-recorded material can make the difference in the quality of the live subtitling produced. For this reason, we concur with the external reviewers' recommendation to make as much material of this type available to subtitlers whenever possible, and in particular for popular entertainment programmes of the type included in our sample.

---

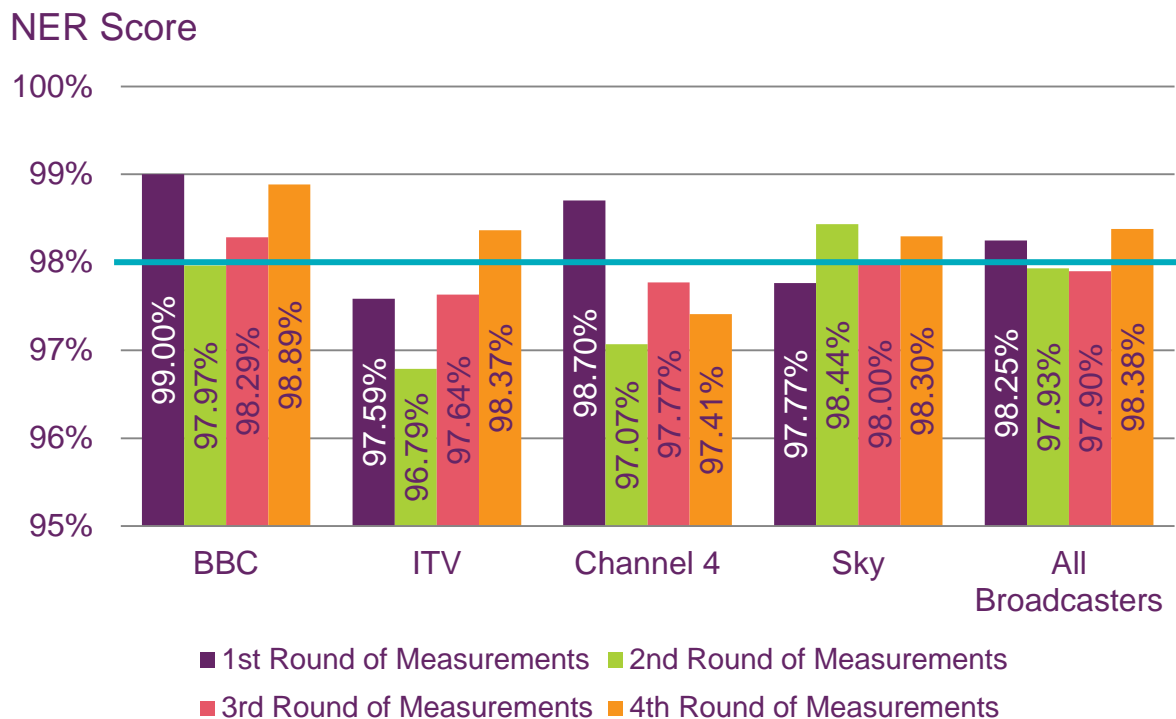
<sup>14</sup> Respectively, the proportion of samples with accuracy rates higher than 98% in previous sampling rounds was: 76%, 74%, and 77%.

<sup>15</sup> Respectively, the proportion of samples with accuracy rates higher than 99% in previous sampling rounds was: 26%, 24%, and 24%.

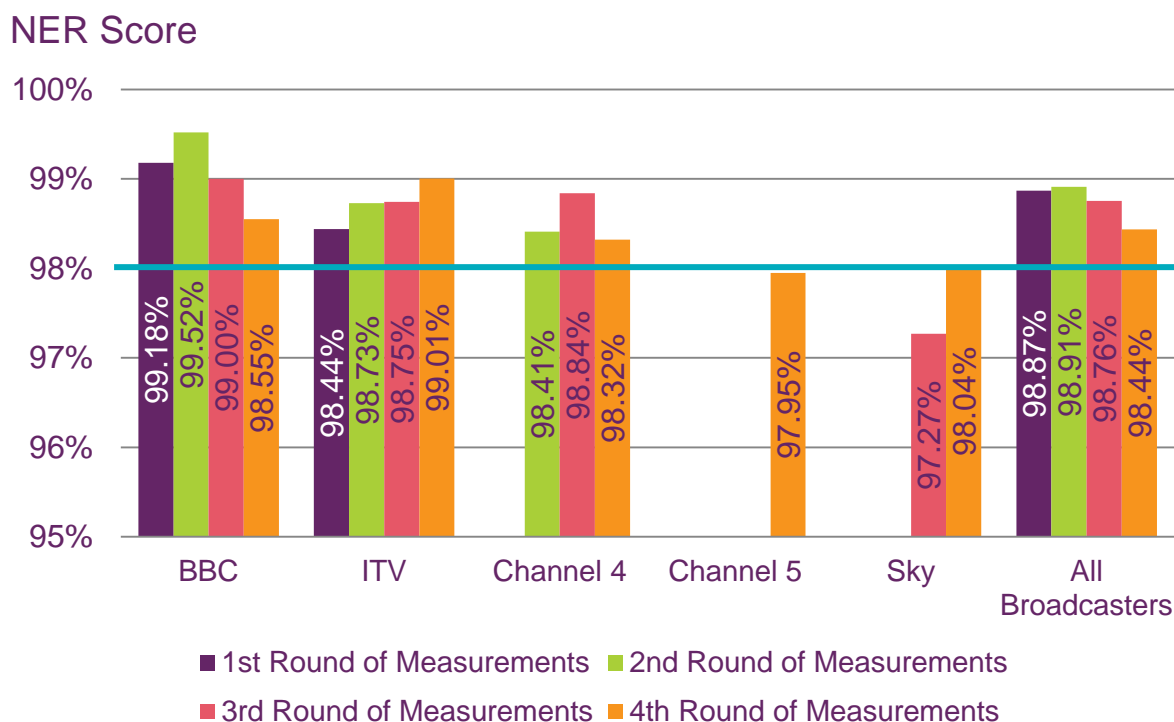
**Figure 1: median accuracy rates in samples drawn from news bulletins**  
(higher bars represent a better viewing experience)



**Figure 2: median accuracy rates in samples drawn from chat shows**  
(higher bars represent a better viewing experience)



**Figure 3: median accuracy rates in samples drawn from entertainment programmes**  
(higher bars represent a better viewing experience)



Note: the first round of measurements did not include samples of entertainment programming shown on Channel 4 and Sky; the first two rounds of measurements did not include samples of entertainment programming shown by Sky; the first three rounds of measurements did not include samples of entertainment programming shown on Channel 5

## Adjusted results

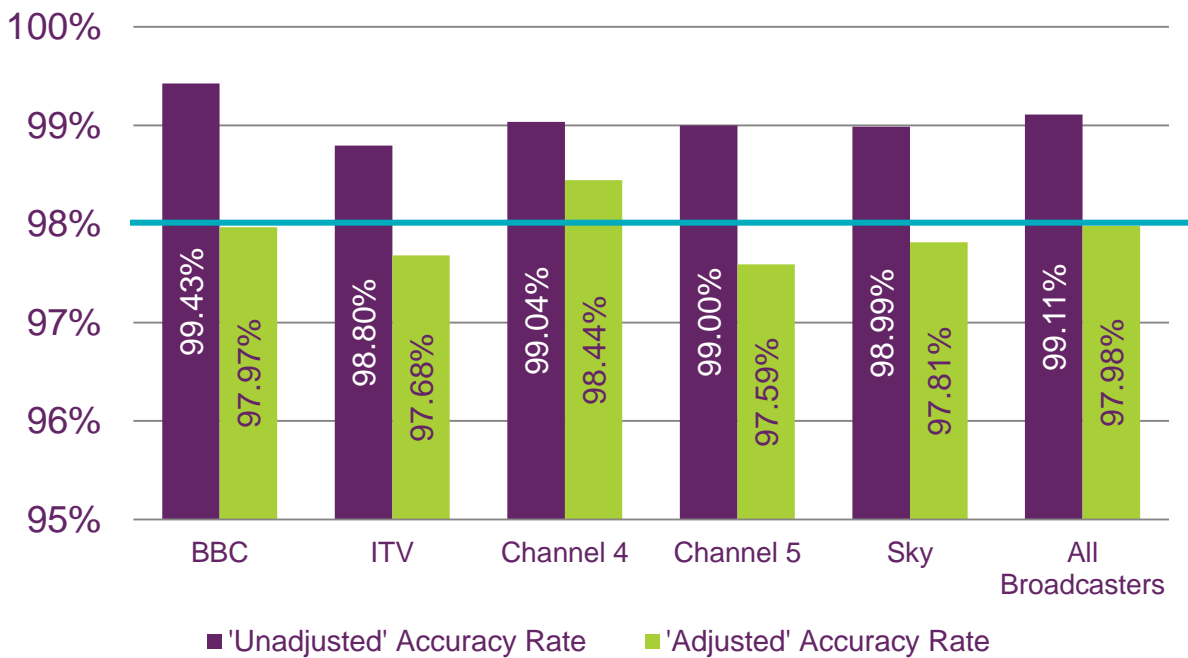
- 2.13 Figures 4, 5 and 6 compare the unadjusted and adjusted accuracy rates in the fourth round of sampling. In this context, every accurate subtitle shown at a speed above 200 wpm has been treated as a standard error. Ofcom recognises that the adjusted results are an imprecise measure of the accuracy of subtitles that include incidences of rapid subtitling. However, we consider that adjusted measurements better represent the quality of the viewing experience of subtitle users than the unadjusted results. Research suggests that many subtitle users will not have enough time to read and comprehend rapid subtitles, and all of them will have much less time to view the images on screen as a result.
- 2.14 Once the median scores are adjusted to account for peaks in subtitling speed, the proportion of samples with accuracy rates above the acceptability threshold falls from 80% to 50%, while the proportion of samples with accuracy rates above 99% falls from 29% to 5%. These results are better than in the third round of measurements, when the proportion of samples with adjusted accuracy above 98% was 32%, and the proportion of very good or excellent adjusted accuracy was less than 3%.
- 2.15 Rapid subtitles frequently occur at the intersection between live scrolling subtitles and pre-recorded block subtitles that were prepared before transmission. For most news bulletins and for a number of entertainment programmes, subtitlers are often provided with pre-recorded material and scripts by the producers of such programmes, so that block subtitles can be prepared in advance and 'keyed out' live. Conversely, the lack of scripts or pre-recorded material means that pre-recorded subtitling is seldom used in chat shows.

2.16 For this reason, it is not surprising that the largest drop in the median accuracy rates measured in the fourth round of sampling – once these are adjusted to account for rapid subtitles – occurred in samples of news bulletins (-1.13 percentage points). As a result, the median adjusted accuracy score for news is 97.98%, just below the level deemed acceptable. By comparison, adjusting the accuracy measurements of samples of chat shows and entertainment programmes resulted in slightly smaller drops (respectively, -0.36 and -0.55 percentage points). The median adjusted accuracy rate remained above 98% for samples of chat shows (98.17%), while it dipped below the acceptability threshold for samples of entertainment programmes (97.83%).

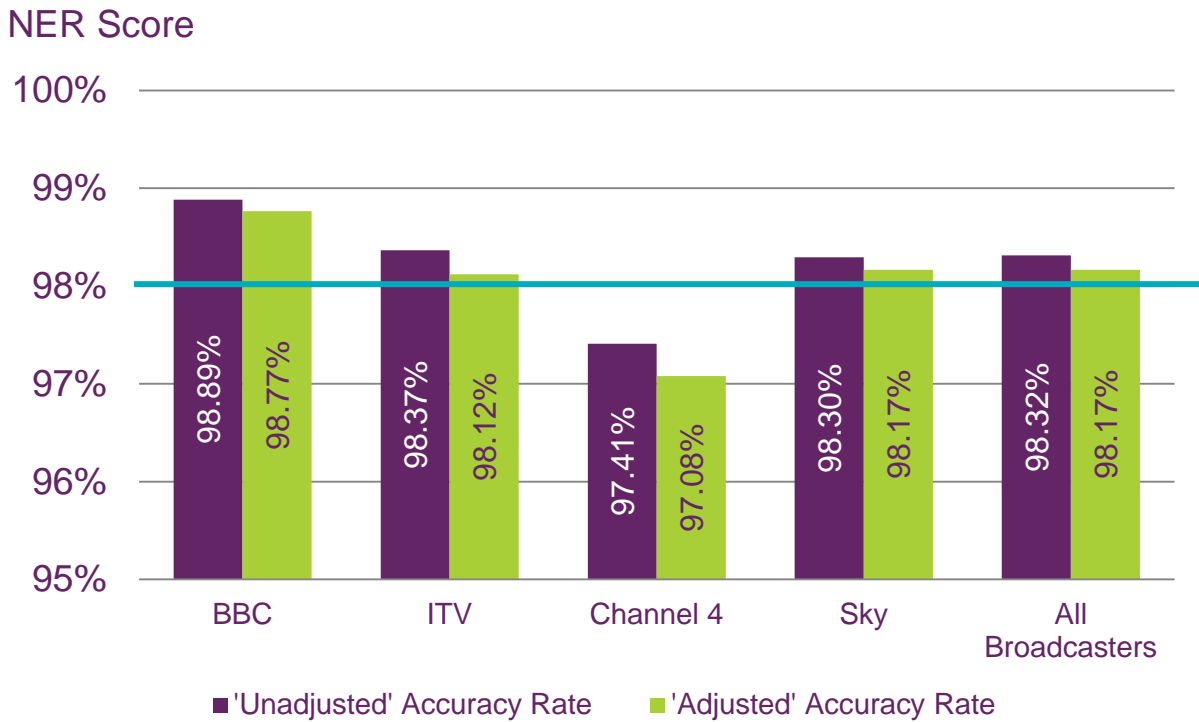
**Figure 4: median accuracy rates in samples drawn from news bulletins (unadjusted vs adjusted to account for rapid subtitles)**

(higher bars represent a better viewing experience)

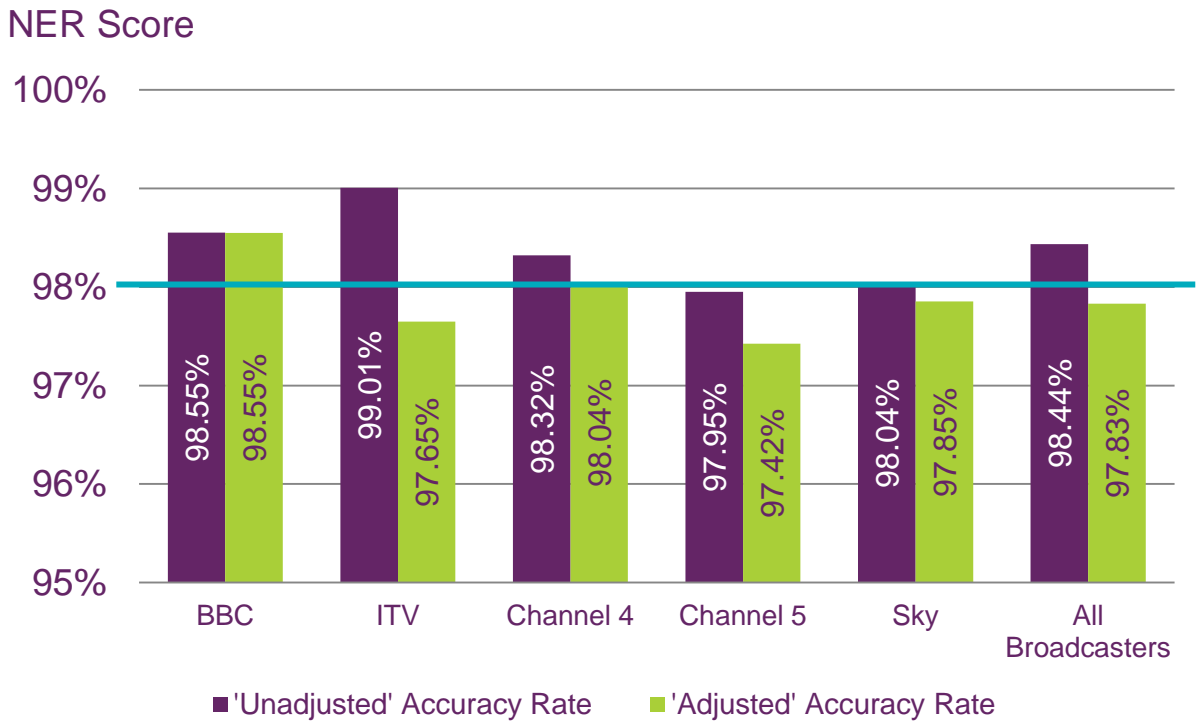
NER Score



**Figure 5: median accuracy rates in chat show samples (unadjusted vs adjusted to account for rapid subtitles)**  
 (higher bars represent a better viewing experience)



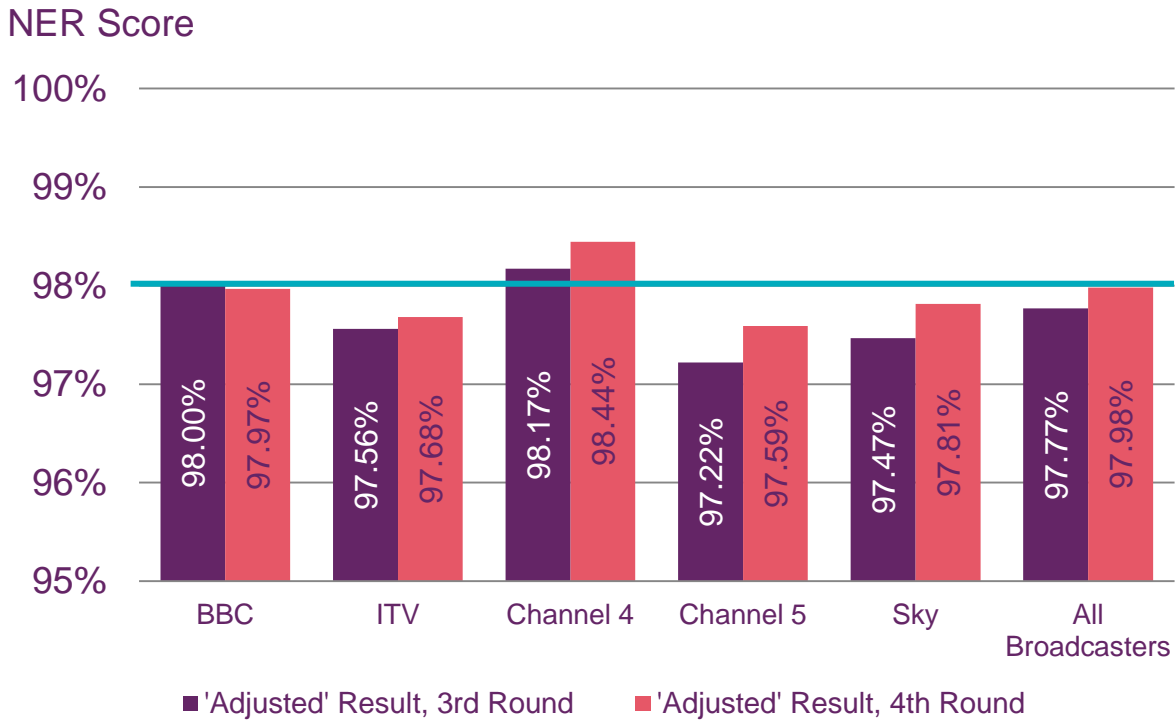
**Figure 6: median accuracy rates in samples drawn from entertainment programmes (unadjusted vs adjusted to take account of rapid subtitles)**  
 (higher bars represent a better viewing experience)



2.17 Figures 7, 8 and 9 compare the adjusted accuracy rates measured in the third and fourth round of sampling. These show that for news bulletins the median adjusted accuracy rates in the fourth round (97.98%) were higher than in the third round (97.77%), although still below the acceptability threshold. The same was true for chat shows (97.64% in the third round, as opposed to 98.17% in the fourth). The median adjusted accuracy rate in samples of entertainment shows was lower in the fourth round (97.83%) than in the third round (98.20%).

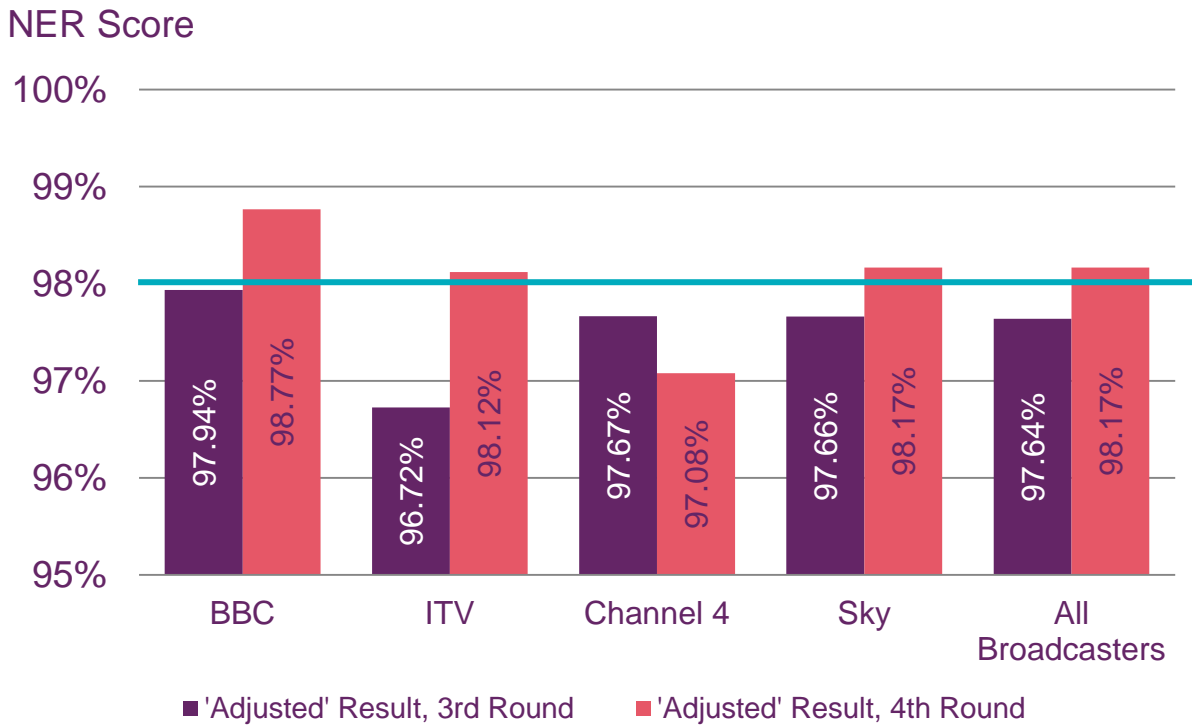
2.18 Data on the incidence of subtitling peaks is provided in Section 4 of this document.

**Figure 7: median adjusted accuracy rates in samples drawn from news bulletins**  
(higher bars represent a better viewing experience)

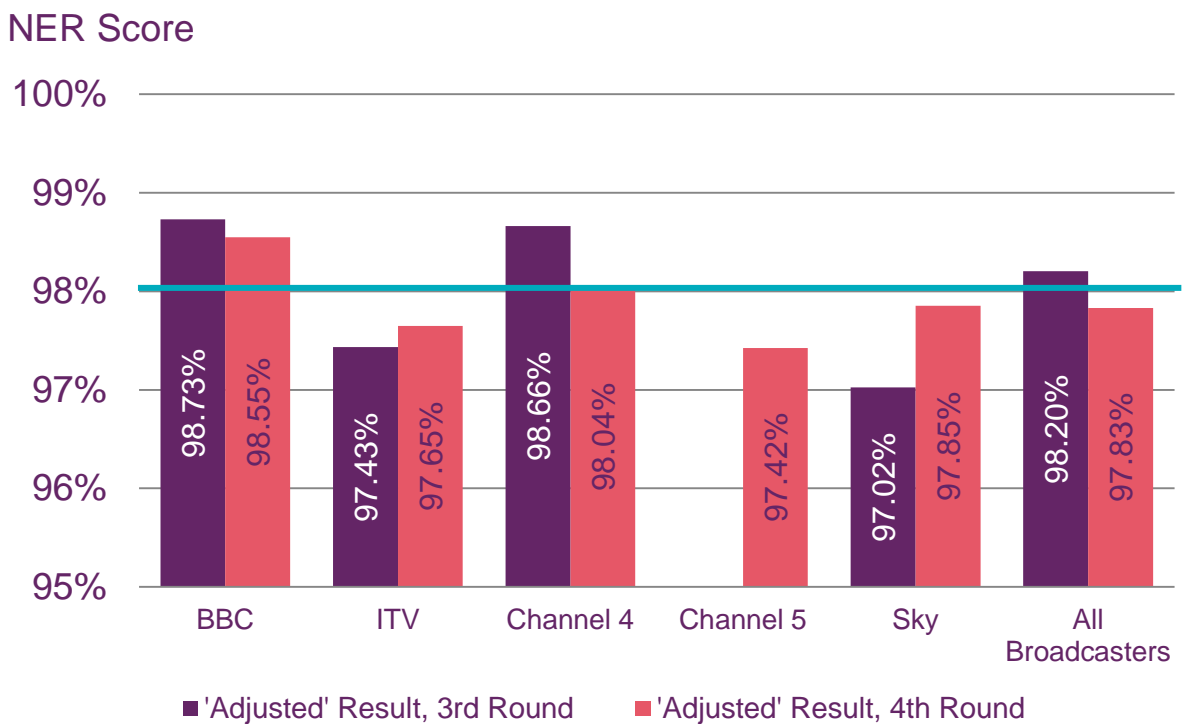




**Figure 8: median adjusted accuracy rates in samples drawn from chat shows**  
 (higher bars represent a better viewing experience)



**Figure 9: median adjusted accuracy rates in samples drawn from entertainment programmes**  
 (higher bars represent a better viewing experience)



Note: the third round of measurements did not include samples of entertainment programming shown on Channel 5

## Section 3

# Latency

- 3.1 In the context of our work on the quality of live subtitling, we refer to 'latency' to denote the delay between speech and the corresponding subtitling. Many subtitle users consider latency as the most frustrating aspect of live subtitling, as it forces them to relate the subtitles to images that may have disappeared from the screen. The longer the delay, the less satisfactory is the viewing experience. Ofcom's Guidelines recommend a maximum latency of 3 seconds for live subtitling<sup>16</sup>.
- 3.2 For the purposes of this monitoring exercise, we have asked broadcasters to identify 2-3 words per minute in the audio at intervals of 20-30 seconds, and to measure the delay between the spoken word and the subtitles. As a result, each 10-minute sample considered for this report had numerous measurements of latency, and these were averaged in order to obtain a unique measure per sample. It should be noted however that the range of latency in most programmes was quite significant.

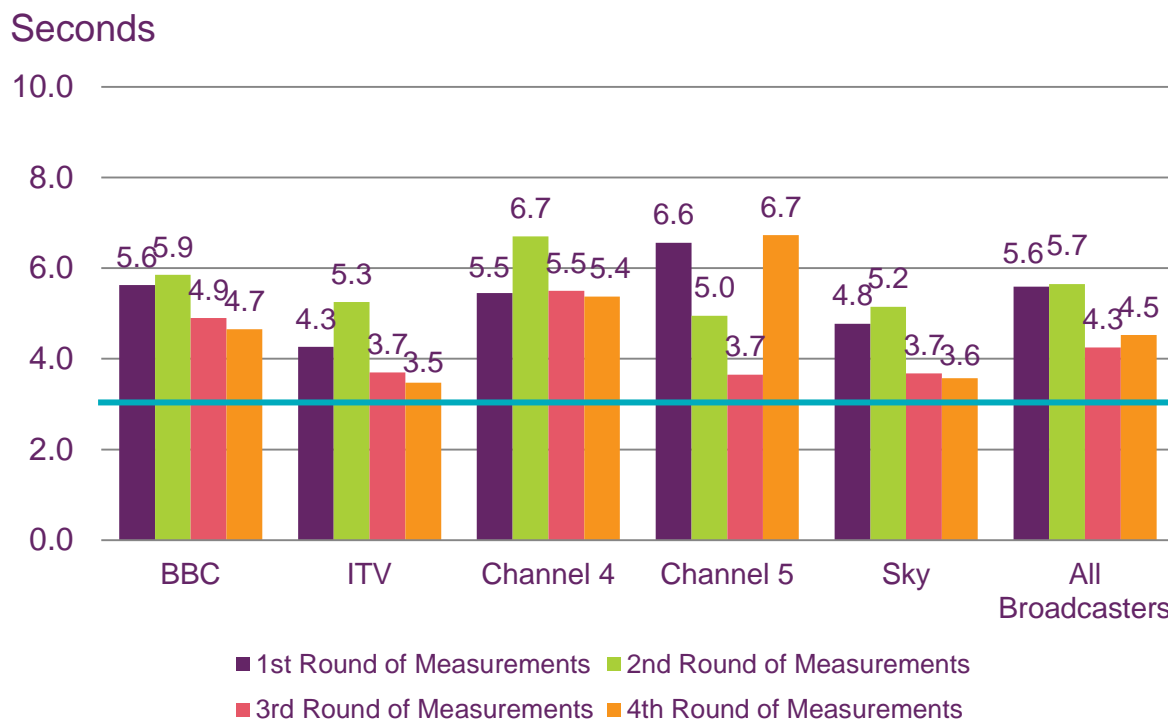
## Summary of results

- 3.3 Figures 10, 11 and 12 below show median latency in the samples taken from news bulletins, chat shows and entertainment programmes. The maximum 3 second delay recommended in Ofcom's guidelines is shown with a bold blue line.
- 3.4 The overall average latency in the fourth round of measurements (5.6 seconds) was broadly in line with the first two reports (5.4s and 5.6s respectively), but higher than the third one (5.1 seconds), and well above the 3 seconds recommended in Ofcom's Guidelines.
- 3.5 In their report, the external reviewers noted that latency typically depends on a number of factors, such as the availability of pre-recorded subtitles, the subtitling techniques used, and specific technical issues during transmissions. Hence, when scripts were made available to subtitlers so they could prepare block subtitles in advance, the average latency was reduced, in a few cases (six) even below the recommended 3 seconds. Conversely, where subtitlers did not have the opportunity to produce block subtitles in advance, and 'key these out' in synchronicity with the original speech – i.e. in cases of 'pure' live subtitling – the average latency was around 7-8 seconds, with peaks of 10 to 21 seconds.
- 3.6 Looking at the individual genres, the median latency in news bulletin samples was slightly higher than in the third round of measurements (4.5 seconds versus 4.3 seconds), although still considerably lower than in the first two sampling rounds (respectively, 5.6 and 5.7 seconds). Chat shows typically feature many speakers, fast speech rates, overlapping voices and no advance scripts; as a result, the median latency on samples of this genre was somewhat higher, at 6.1 seconds, and broadly in line with previous sampling rounds. The median latency recorded in samples of entertainment programmes was the highest since the start of the monitoring exercise, at 6.1 seconds. This was partly due to the increased proportion in our sample of entertainment programmes with characteristics similar to chat shows (see paragraph 2.9).

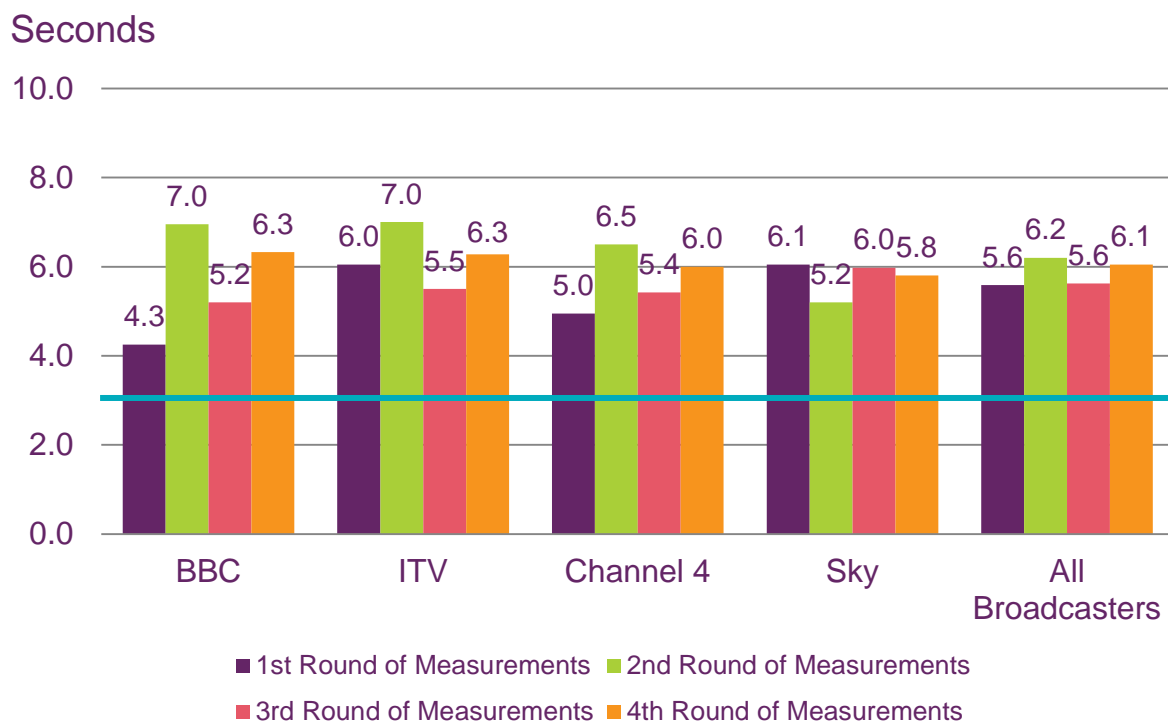
---

<sup>16</sup> Ofcom's Code on Television Access Services, 15 May 2015; Annex 4, paragraph A4.18 (<http://stakeholders.ofcom.org.uk/broadcasting/broadcast-codes/tv-access-services/code-tv-access-services-2015/>).

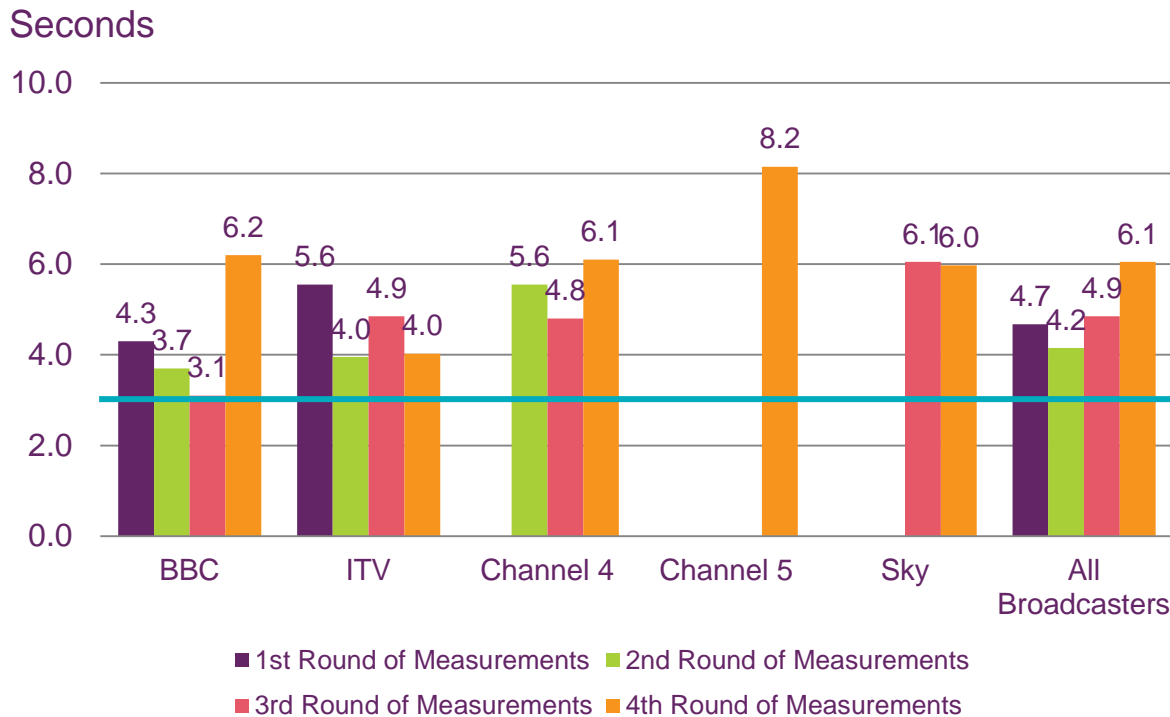
**Figure 10: median latency in samples drawn from news bulletins**  
 (higher bars represent an inferior viewing experience)



**Figure 11: median latency in samples drawn from chat shows**  
 (higher bars represent an inferior viewing experience)



**Figure 12: median latency in samples drawn from entertainment programmes**  
(higher bars represent an inferior viewing experience)



Note: the first round of measurements did not include samples of entertainment programming shown on Channel 4 and Sky; the first two rounds of measurements did not include samples of entertainment programming shown by Sky; the first three rounds of measurements did not include samples of entertainment programming shown on Channel 5

## Tackling latency

- 3.7 Since our 2013 consultation<sup>17</sup>, we have had extensive discussions with broadcasters about whether the intractable problems of latency could be addressed by inserting short delays in the transmission of certain programmes that are subtitled live (excluding news and sports programming). While broadcasters argued that the cost and complexity of inserting short delays would be disproportionate to the benefits that subtitle users would enjoy, the BBC and Channel 4, working with their access service provider, Ericsson (formerly Red Bee Media), have identified an alternative approach that could reduce latency significantly – so-called ‘switchable delays’.
- 3.8 Specifically, this technical solution for latency takes advantage of the time it takes for the play-out system to encode a High Definition (‘HD’) picture ready for transmission (approximately 5 seconds). Encoding the data comprising the subtitles takes a much shorter time. At present, broadcasters therefore insert a short delay in the transmission of subtitling, in order for it to be synchronised with the speech. In order to address the problem of latency in live subtitling, the BBC has experimented with ‘re-timing’ the subtitle files on live-subtitled programmes separately from the HD picture to better align the subtitles with the picture and sound. Ofcom understands that the trials conducted in the BBC R&D lab have yielded some promising results. Channel 4 has told us that it is planning a proof-of-concept trial of a very similar approach. Both have agreed to share their findings with interested parties. We have invited ITV, Sky and Channel 5 to explore the potential of this approach, but we

<sup>17</sup> Ofcom, *The quality of live subtitling – Improving the viewer experience*, 17 May 2013

recognise that broadcasters use different technical systems and that this approach may not work for all.

- 3.9 We understand that the timing of any implementation would be dependent in part on channels being equipped with up to date equipment, and that roll-out could take some years. In particular, the BBC has told us that this approach would be available only when its new playout system is rolled out on BBC 4, CBBC and CBeebies in early 2016, while it will be rolled out on BBC One and Two during 2017. Initially, the results will only be available on HD channels, though there is the prospect of making it available on SD channels later.
- 3.10 However, the reduction in latency resulting from these plans would not tackle the particular problem of poorer accuracy for some popular entertainment and chat shows. Subtitling on these is also typically more edited than on other programmes (see Section 5), something that some subtitle users consider problematic. Some of these programmes, like *Gogglebox*, are pre-recorded, and Ofcom believes that broadcasters should make greater efforts to enable subtitles for such programmes to be prepared in advance, even if they have to be cued out live.
- 3.11 For the time being, we remain unpersuaded that short delays in live transmissions would not be technically feasible. The Flemish public service broadcaster, Vlaamse Radio- en Televisieomroeporganisatie ('VRT'), has been introducing delays in the transmission of some of its programmes in order to improve the quality of the live subtitling for a number of years now<sup>18</sup>. Last month, VRT demonstrated to Ofcom and others how it uses short delays to produce higher quality subtitles. It now produces about 20 hours a week of programmes with 'delayed live subtitling'. VRT uses three subtitlers: one to produce the subtitles, one to edit them for accuracy, and one to cue them out at the correct time. While more expensive than the existing approach, this level of resourcing would only be required for relatively few programmes. However, we consider that it is likely to be more fruitful in the medium term for broadcasters to concentrate their efforts on using switchable delays to secure significant improvements in latency, where their technology allows.

---

<sup>18</sup> See: Ofcom, *Measuring live subtitling quality: results from the third sampling exercise*, 20 May 2015 ([http://stakeholders.ofcom.org.uk/binaries/consultations/subtitling/statement/QoS\\_3rd\\_Report.pdf](http://stakeholders.ofcom.org.uk/binaries/consultations/subtitling/statement/QoS_3rd_Report.pdf)); page 15.

## Section 4

# Speed of subtitling

- 4.1 Ofcom’s guidelines recommend that “the speed should not normally exceed 160 to 180 wpm for pre-recorded programmes”<sup>19</sup>; however, there is no recommended maximum speed for live subtitling.
- 4.2 Research<sup>20</sup> suggests that the speed of subtitling has a direct impact both on subtitle users’ ability to comprehend the subtitles, and on the amount of time viewers are able to watch what is happening on screen, as opposed to reading subtitles. Furthermore, people cannot typically read as fast as they can speak. At a rate of more than 180 wpm, deaf, hard-of-hearing and hearing readers found it difficult, and the optimum rate has been found to be lower than that<sup>21</sup>.
- 4.3 As discussed in Section 6, our review of the Guidelines on the provision of access services which will follow the publication of this report will take account of the views of external stakeholders, as well as lessons from the research we plan to commission, and other recent research.
- 4.4 Over the course of our work on the quality of live subtitling, there has been a visible increase in the use of block subtitles for scripted or pre-recorded segments in live programmes. These are usually prepared in advance, thanks to relevant material being provided to subtitlers in before transmission. From the second report, the external reviewers started noting that, at the junction between scrolling subtitles (always delayed) and pre-prepared subtitles (potentially synchronous), some subtitlers attempted to synchronise the pre-recorded subtitles to the audio as quickly as possible, when these followed a segment with naturally delayed live subtitles.<sup>22</sup> As a result, some of the block subtitles at these intersections were shown at unreadable speeds.
- 4.5 In order to understand the impact of this practice, we decided to look in more detail at the issue of very fast subtitles in the third sample. For this purpose, we asked broadcasters to provide us with time-coded transcripts of both the original audio and the subtitles in order to identify all the ‘rapid subtitles’ (those faster than 200 wpm). We asked broadcasters to do the same for the purpose of the fourth report.
- 4.6 Below, we report first on average speeds (as we have from the beginning of the monitoring exercise), and then on the incidence of rapid subtitles in the fourth round of measurements.

## Average speeds

- 4.7 The overall average subtitling speed measured in the fourth sampling round was 133 wpm, well below Ofcom’s guidelines as in the case of the previous reports. Only one sample had an average speed above 180 wpm, and only less than 10% of the

<sup>19</sup> Ofcom’s Code on Television Access Services, 15 May 2015; Annex 4, paragraph A4.19

<sup>20</sup> Among others: Jensema, Carl: *Viewer Reaction to Different Captioned Television Speeds*, June 1997 (<https://www.dcmp.org/caai/nadh30.pdf>); Romero-Fresco, Pablo: *More haste less speed: Edited versus verbatim respoken subtitles* (<http://vialjournal.webs.uvigo.es/pdf/Vial-2009-Article6.pdf>)

<sup>21</sup> *ibid*

<sup>22</sup> Ofcom, *Measuring live subtitling quality – Results from the second sampling exercise*; paragraph 4.8.

samples showed an average subtitling speed higher than 160 wpm. The median average speed was highest during news bulletins (148.8 wpm) than in chat shows (127.1 wpm) and entertainment programmes (117.9 wpm), due to the higher speech rates and to comparatively low edition rates (i.e. the amount of original speech that is left out of the subtitles - see Section 5).

- 4.8 Figures 13, 14 and 15 below show the median wpm rates in the samples taken from news bulletins, chat shows and entertainment programmes. For reference purposes, the upper (180 wpm) and lower (160 wpm) boundaries of the maximum range recommended for pre-recorded subtitling is indicated on the graphs below with bold coloured lines.

**Figure 13: median wpm rates in samples drawn from news bulletins**

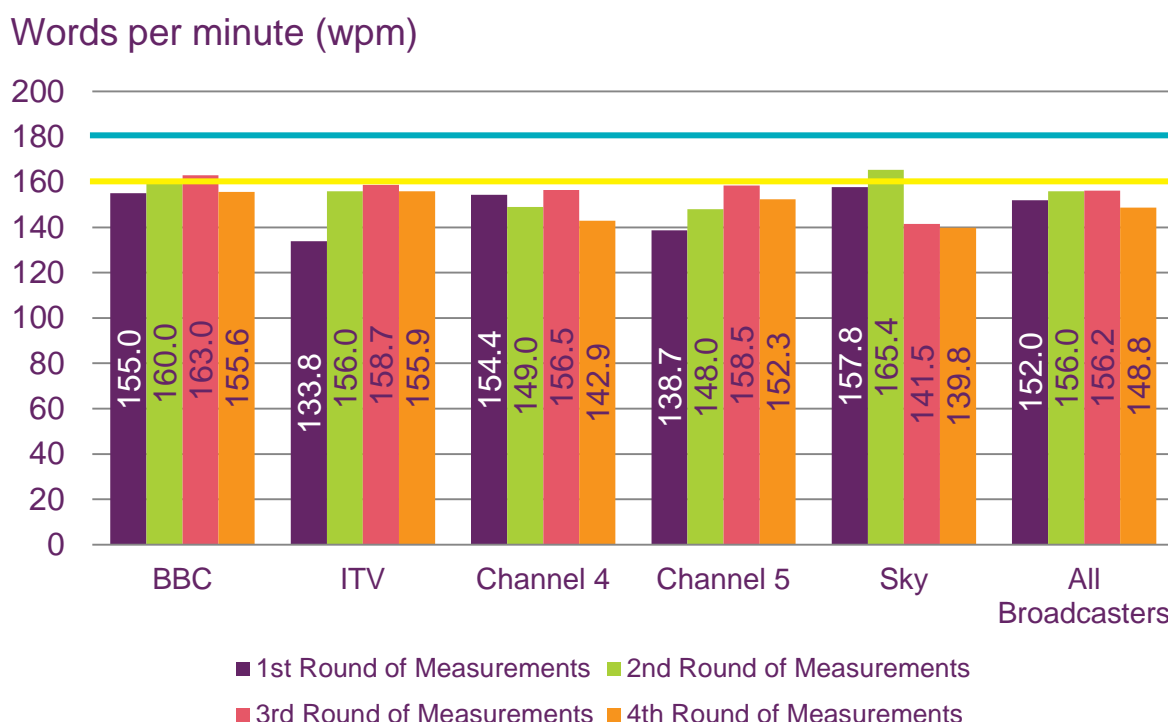


Figure 14: median wpm rates in samples drawn from chat shows

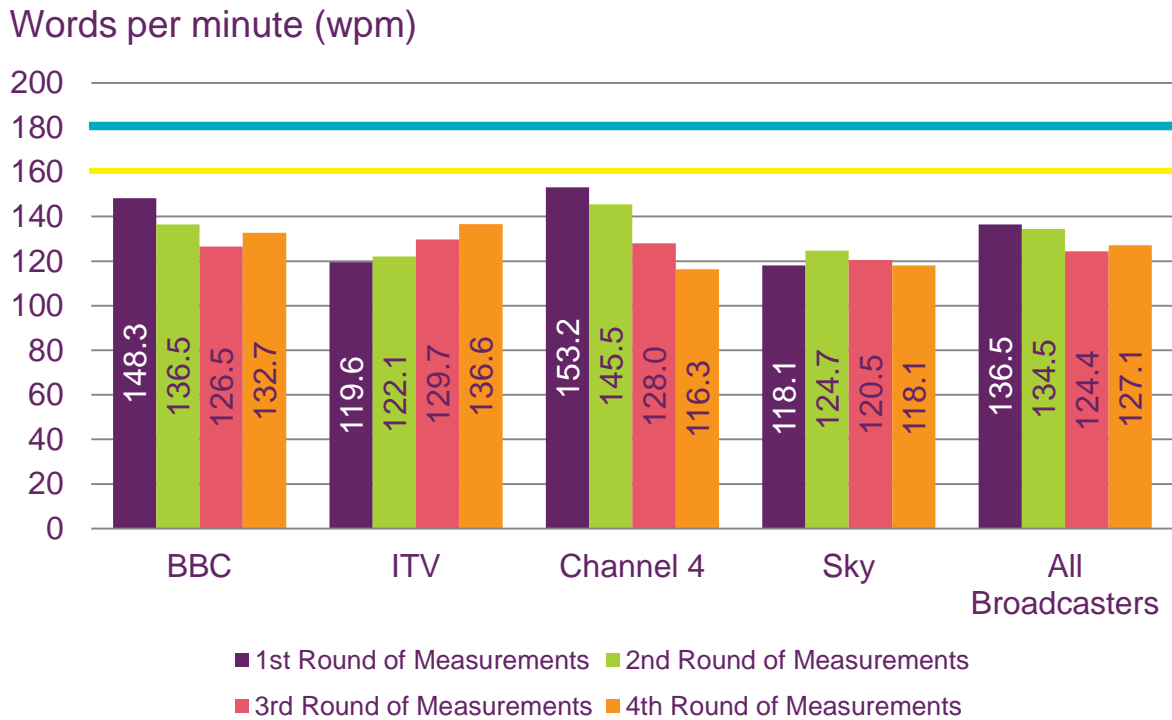
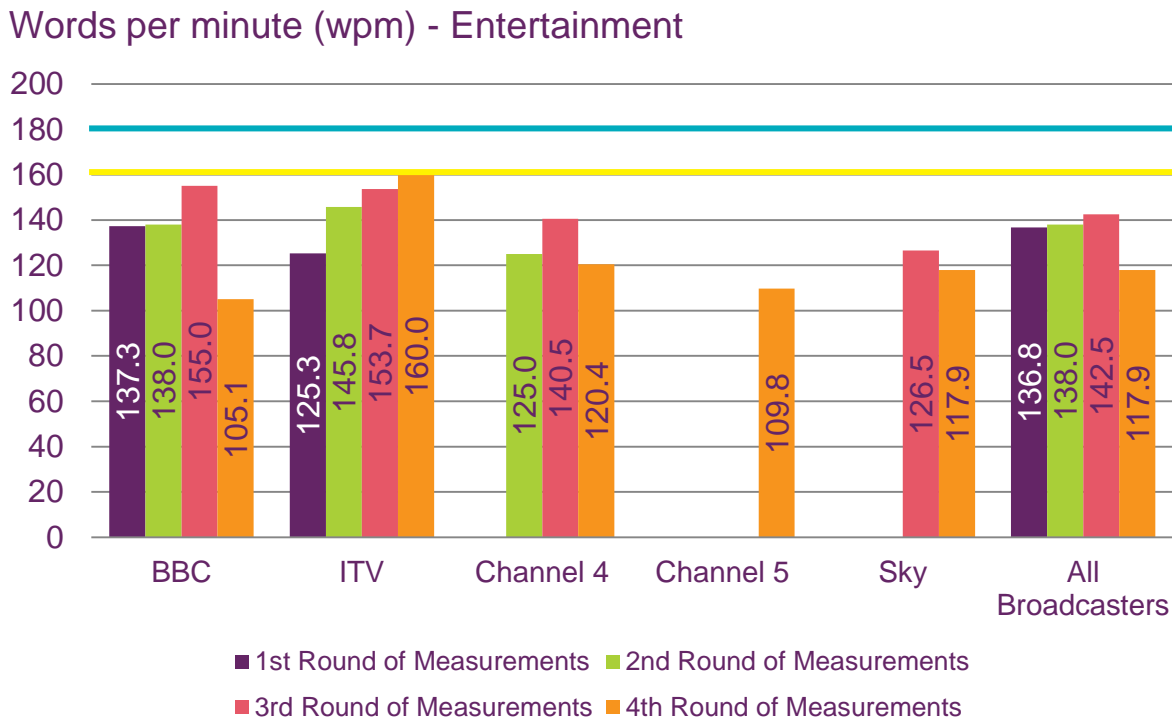


Figure 15: median wpm rates in samples drawn from entertainment programmes



Note: the first round of measurements did not include samples of entertainment programming shown on Channel 4 and Sky; the first two rounds of measurements did not include samples of entertainment programming shown by Sky; the first three rounds of measurements did not include samples of entertainment programming shown on Channel 5

### Incidences of rapid subtitling

4.9 The aggregate speed measures reported above fail to illustrate the potential impact of sudden spikes in the speed of subtitles on the viewer experience. It appears that



instances of rapid subtitles have become more common as 'hybrid subtitling' – i.e. subtitling which combines pre-recorded, block subtitles and scrolling, live subtitles – has become commonplace on certain types of programmes, news in particular. In order to measure the extent of rapid subtitling, we asked broadcasters to provide us with time-coded transcripts of both the original audio and the subtitles, so that the external reviewers could identify all the instances where subtitling speed exceeded 200 wpm.

- 4.10 Figures 16, 17 and 18 show the incidence of rapid subtitles in all the samples taken from news bulletins, chat shows and entertainment programmes. Slightly more than 90% of the samples analysed for the fourth round of measurements included rapid subtitles, with an average of 21 rapid subtitles per sample, and a median of 12. In the third round of measurements, rapid subtitles were found in 99% of the samples, and the average number of rapid subtitles per sample was 26, while the median was 21.
- 4.11 Pre-recorded subtitles are used extensively for news bulletins, for which subtitlers are provided with running orders, scripts and pre-recorded material. They are also used for some entertainment programmes, particularly those with clearly pre-defined formats or with pre-recorded videos. Pre-recorded subtitling is seldom used in chat shows, as the conversation on these programmes tends not to follow a script. As discussed above, rapid subtitles feature more prominently at the intersection between pre-recorded and live subtitling.
- 4.12 So it is not surprising to find that the samples of chat shows were less affected by the problem of very fast subtitling, with an average of 6 rapid subtitles per sample, compared to an average of 11 in samples of chat shows from the third round of measurements. Rapid subtitles were more common in samples of news programmes, with an average of 35 instances of very fast subtitling per sample, slightly fewer than in the previous report (40). Samples of entertainment programmes were not as affected by rapid subtitles (an average of 14 per sample), with the exception of programmes shown on ITV, which featured more pre-recorded subtitles.

Figure 16: number of times in each 10-minute sample when subtitles exceeded 200 wpm, all samples (news bulletins)

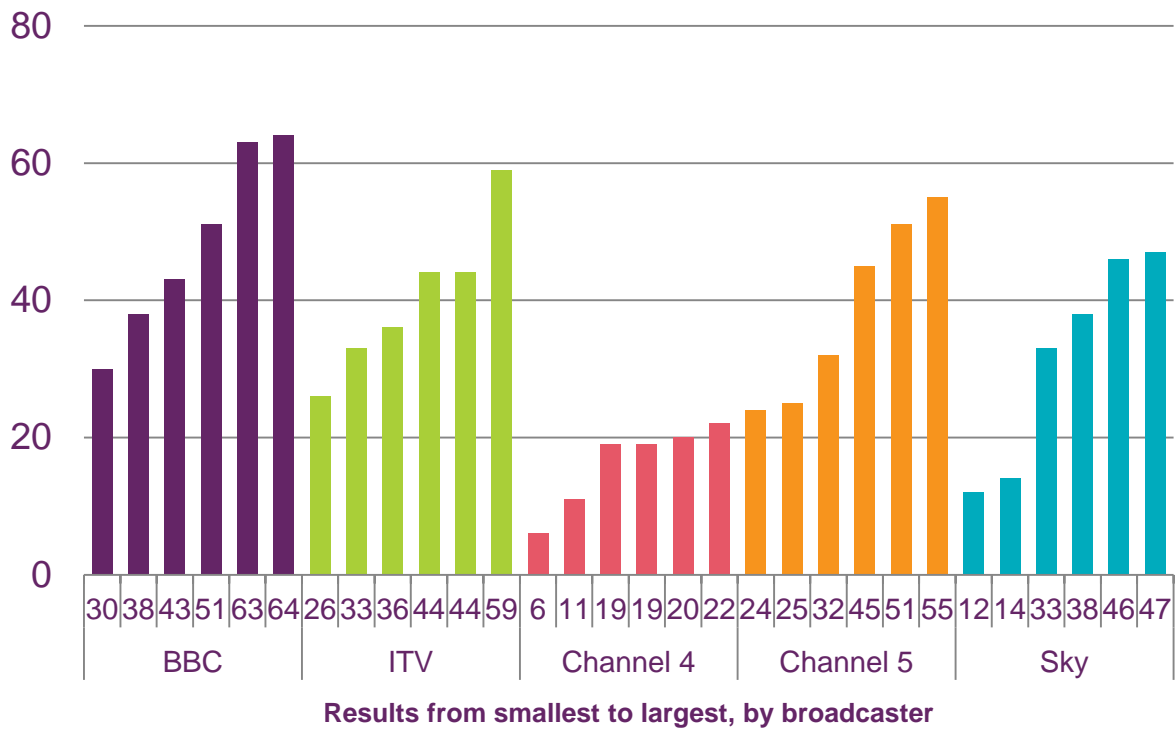
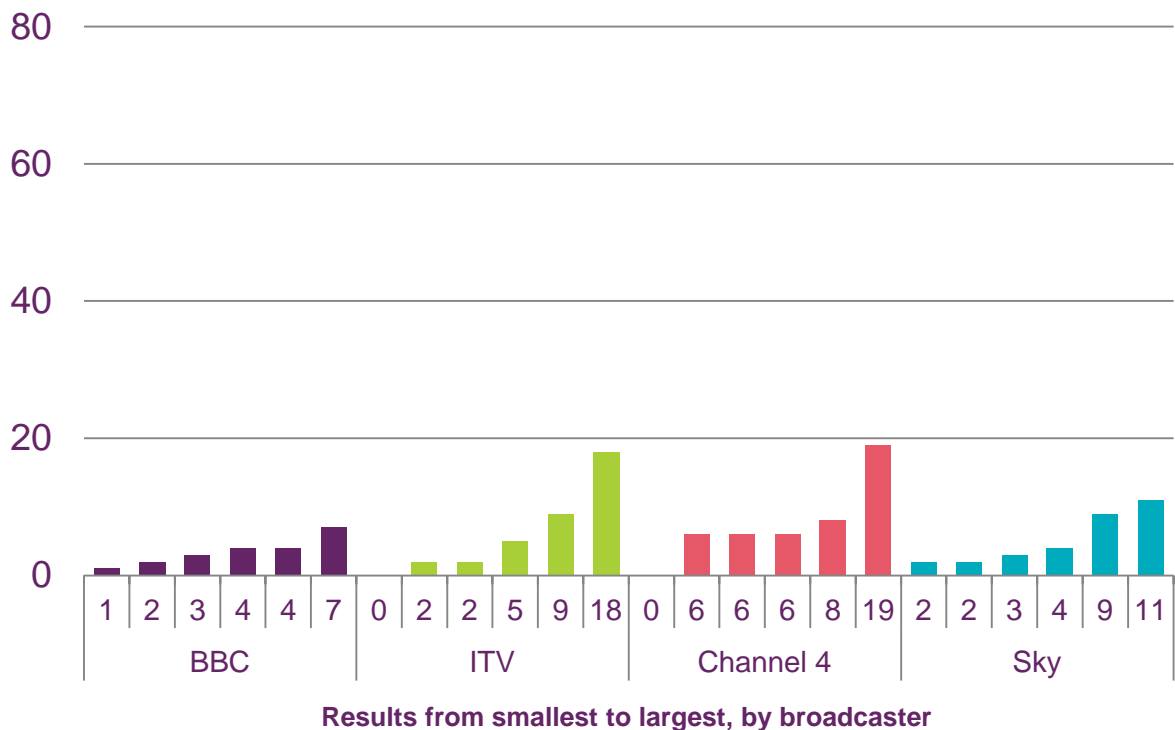
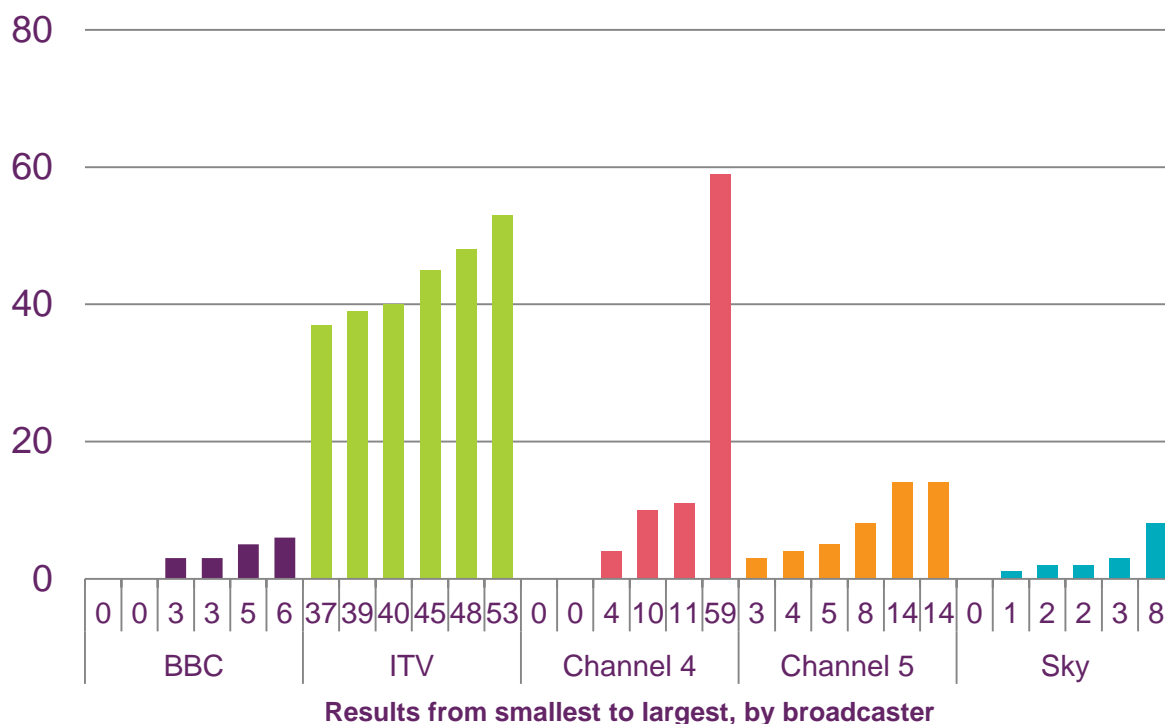


Figure 17: number of times in each 10 minute sample when subtitles exceeded 200 wpm, all samples (chat shows)



**Figure 18: number of times in each 10 minute sample when subtitles exceeded 200 wpm, all samples (entertainment programmes)**



### Mechanisms for setting limits to the speed of subtitling

4.12 Ofcom welcomes the contribution that hybrid subtitling could make to increased accuracy; however, the data on rapid subtitles above, together with the data on adjusted accuracy rates reported in Section 2, indicates that there is a trade-off to be made between greater accuracy and slower, more readable subtitles. We consider this trade-off in greater detail in Section 6. In our previous report we encouraged broadcasters to address the problems that rapid subtitling created, and we said that we would ask broadcasters to report on what measures they were taking or planning to take to avoid rapid subtitling. We noted that the external reviewers from the University of Roehampton recommend “that the subtitling software is set to a maximum speed (ideally no higher than 180 wpm-200 wpm) for both pre-recorded and live scrolling subtitles”.

4.13 Below is a summary of the broadcasters’ responses:

- a) the BBC agreed with Ofcom’s assessment that incidence of rapid subtitling can result in the degradation of the viewing experience. However, the BBC also noted that there are programmes where the participants will talk at rates over 200 wpm, and that for this reason a maximum speed limit of 200 wpm might increase latency as well as edition rates. According to the BBC, more research in this area is required.

The BBC told Ofcom that it is currently working with its subtitling provider to identify the causes of high subtitle reading rate in order provider to improve the practices when live and block segments intersect, and it is also working to improve its network and broadcast infrastructure to reduce the impact of ‘bottlenecks’ – i.e. congestion in the delivery of subtitling due to a technical fault, resulting in subtitles

appearing for only fractions of a second on the screen – for the live subtitle data stream;

- b) ITV also agreed with Ofcom’s assessment that rapid subtitling is a problem that requires further attention. ITV’s subtitling provider has now updated its subtitling software in order to allow the setting of a maximum reading speed. This feature ensures that block subtitles remain on screen for a minimum length of time. The subtitling provider has experimented with various configurations of this limit, and it has noticed that there is a negative correlation between the maximum speed limit and the increase in latency of the subtitles. In the light of this, ITV’s subtitling provider told us that it has set the maximum speed limit set at 320 wpm for the time being;
  - c) Channel 4 noted that its subtitling provider operates maximum speed settings available through their subtitling software. The current speed limit is 250 wpm for block pre-recorded subtitles; we understand from Channel 4’s subtitling provider that there is also a limit of 300 wpm for scrolling live subtitles. However, Channel 4 noted that there are many examples where the real-world live speech on a programme will be above 200 wpm. Channel 4 considers that the aim of broadcasters should be to provide subtitling that is as accurate and relevant to the pictures on screen as possible, and that a 200 wpm speed limit would undermine their ability to achieve this;
  - d) Channel 5 noted that the problem of rapid subtitling is a consequence of the move towards hybrid subtitling, with block and scrolling subtitles both used in a programme. Channel 5’s subtitling provider has tested various speed limits, after which it has implemented a speed limit of 300 wpm which, in its view, addresses the incidence of very high speed subtitles without increasing latency; and
  - e) Sky also noted that the increased use of block pre-recorded subtitling is of the reason for the increased occurrence of rapid subtitles. Sky’s subtitling provider currently implements a limit to subtitling speed at 250 wpm, and it is also providing additional training to subtitlers to ensure that they do not ‘cue’ the block subtitles too quickly following sections subtitled live.
- 4.14 Ofcom recognises that there is a trade-off between latency and speed, and considers that the views of subtitle users should be sought on how best to balance the two factors. We consider that the research we plan to undertake should help to shed light on this (see Section 6).

## Section 5

# Edition rates

- 5.1 Edition rates refer to the proportion of the original speech that is left out (i.e. 'edited out') from the subtitles. In order to keep up with the pace of the original speech and to reduce subtitling speed, re-speakers commonly edit down the text that is subtitled. For the purpose of our reports, edition rates were calculated by dividing the total number of words in the subtitling by the total number of words in the transcript of the original speech from the same 10-minute fully-spoken segment.
- 5.2 In the context of our work on the quality of live subtitling, we have never referred to edition as a dimension of the quality of live subtitling. Sometimes the same idea can be expressed with fewer or different words. So correct editions are not counted against the measure of accuracy defined earlier in this document.
- 5.3 Nevertheless, we consider that the information on how much speech is edited out will help to inform the debate between those seeking near verbatim subtitling, even at the cost of faster subtitling, and those who advocate greater amounts of editing in order to keep the speed of subtitling down.

## Summary of results

- 5.4 Figures 19, 20 and 21 summarise the median edition rates by genre and broadcaster.
- 5.5 Edition rates differ significantly across genres. In samples of news programmes, the median edition rate was relatively low in all sampling rounds, between 12% and 17%. This is due to two main factors:
- a) subtitlers seem to make a significant effort to avoid excessive editing, given the importance of accurate content in this genre;
  - b) the more extensive use of block pre-recorded subtitles, where subtitlers can condense more words, and which do not suffer from lags or errors.
- 5.6 Conversely, on chat shows and on some entertainment programmes, the level of edition is typically higher than on news. In these types of programming, edition is often necessary due to fast speech rates, overlapping speech and quick changes of subject. As a result, in the fourth round of measurements, the edition rate in samples chat shows was almost twice as high as in samples of news, at 27.5%, although slightly lower than in the third round (42.8%). The edition rate in samples of entertainment programmes was the highest since the beginning of the project (32.1%), partly due to a greater proportion of entertainment programmes with similar characteristics to chat shows being included in our sample.
- 5.7 As part of the research into subtitle users' attitudes towards live subtitling that we plan to commission, we also intend to seek an understanding of subtitle users' attitudes towards edition rates.

Figure 19: median edition rate in samples drawn from news bulletins

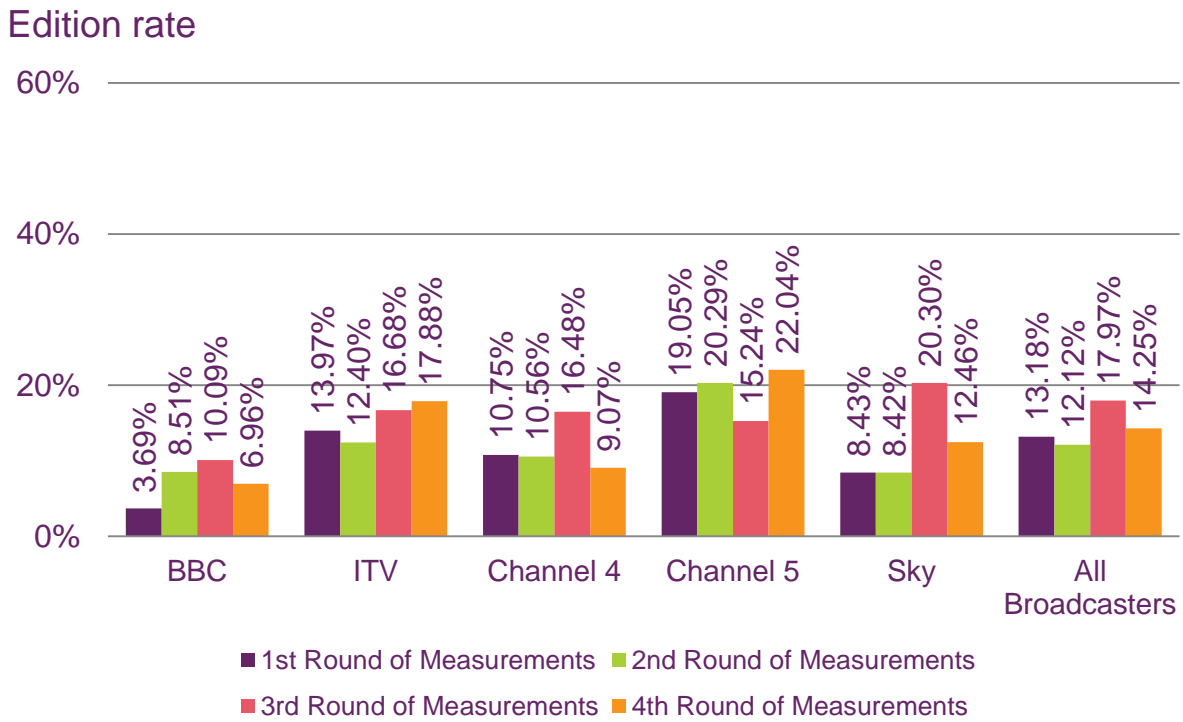
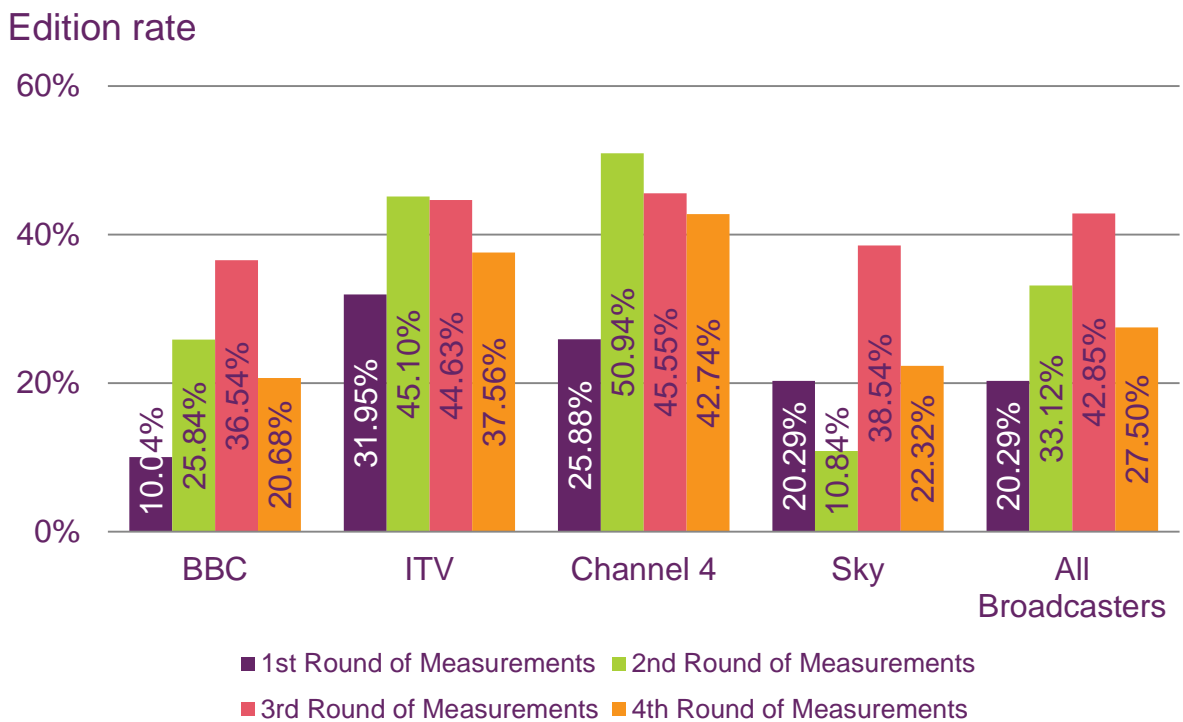
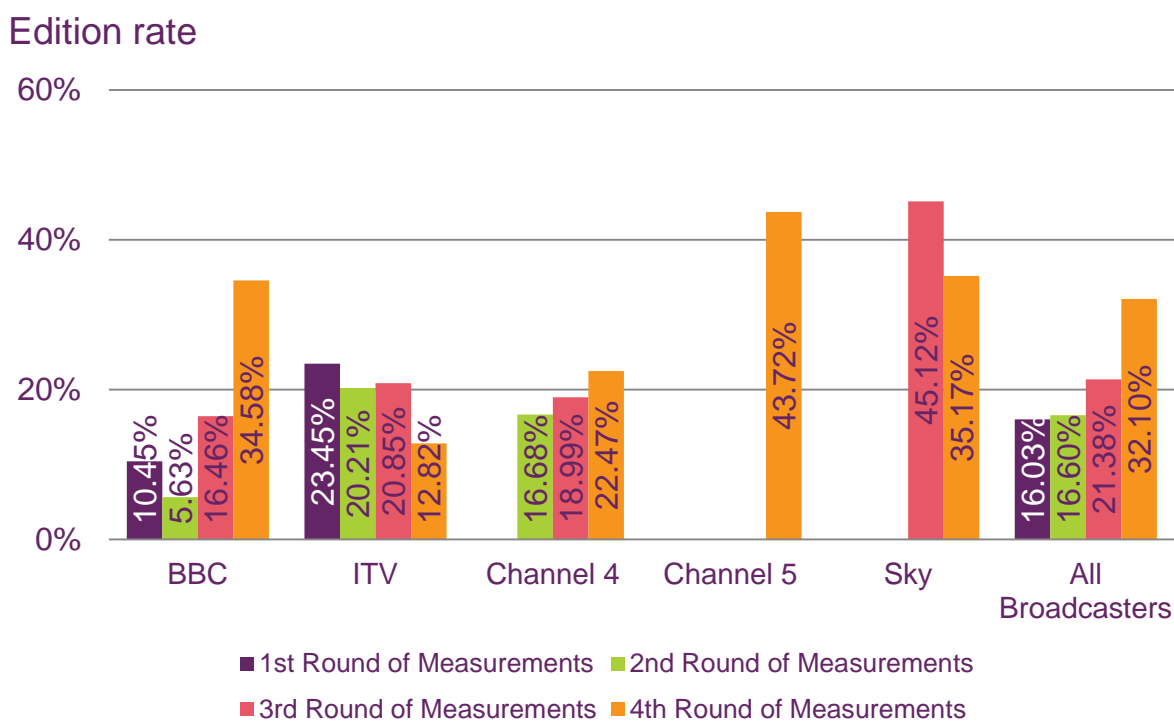


Figure 20: median edition rate in samples drawn from chat shows



**Figure 21: median edition rate in samples drawn from entertainment programmes**



Note: the first round of measurements did not include samples of entertainment programming shown on Channel 4 and Sky; the first two rounds of measurements did not include samples of entertainment programming shown by Sky; the first three rounds of measurements did not include samples of entertainment programming shown on Channel 5

## Section 6

# Next steps

## Introduction

6.1 The evidence gathered from the exercise to measure the quality of the live subtitling will need to be considered carefully alongside available research, and the views of subtitle users, broadcasters and access service providers. Ofcom aims to do this over the next few months. We set out the process we expect to follow below, under 'Next steps'. In the meantime, however, we consider that it would be helpful to publish some of Ofcom's preliminary conclusions, in order that they can inform debate.

## Preliminary conclusions

6.2 We are keen to avoid drawing premature conclusions about the lessons to be learnt from this exercise, as we want to discuss the overall results with subtitle users and groups representing their interests, as well as broadcasters and access service providers. We also want to consider what additional lessons there are from research carried out in recent years.

6.3 Nevertheless, we do believe that some outcomes are clear:

- a) by making every effort to prepare subtitles in advance for pre-recorded programmes, broadcasters and access service providers can make a big contribution to the quality of the viewing experience for subtitle users. Ofcom is encouraged that broadcasters have sought to reduce the incidence of live subtitling for pre-recorded programmes, and will ask them to monitor the extent to which such programmes have to be subtitled live, so that they can report on this from time to time. We hope that programme producers will participate in this monitoring exercise;
- b) by helping access service providers prepare for live subtitling through the provision of running orders, autocue scripts and access to the newsroom, broadcasters have helped to improve the quality of live subtitling for news programmes;
- c) technical problems occur along the subtitling supply chain relatively frequently, although they affect a very small proportion of output. Ofcom will consider how best it can encourage broadcasters to minimise the number of technical problems.

6.4 But our main preliminary conclusion is that, when producing live subtitling, it is likely that trade-offs will need to be struck between different quality dimensions. For example, it would be possible to produce accurate, verbatim subtitles for a fast-paced programme, but only if viewers accepted significant delays in the subtitles appearing.

## We'd like to find out what subtitle users think about likely trade-offs

6.5 While the necessity for trade-offs should always be tested, Ofcom considers that it would be helpful to explore the likely trade-offs with subtitle users, and to ensure that, so far as practicable, the balance struck by broadcasters and access service providers takes account of the preferences of subtitle users. With this in mind, Ofcom



intends to commission independent qualitative research amongst subtitle users on their attitudes to trade-offs.

- 6.6 We have invited the views of groups representing subtitle users, broadcasters and access service providers on our proposals to look at the issues described below.

### **Presentation of subtitles**

- 6.7 Research suggests that viewers find block subtitles easier to read than scrolling subtitles, leaving more time to absorb visual information<sup>23</sup>. There is now more scope for programme subtitling software to format subtitles in blocks automatically, but doing this rather than using scrolling subtitles may result in delays in subtitles appearing on screen. We would like to know which approach would suit most subtitle users.

### **Subtitling speeds in programmes with both live and pre-recorded subtitling**

- 6.8 Live programmes often include pre-recorded segments. Broadcasters now often prepare subtitling in advance for these segments, offering better accuracy, presentation in easier-to-read subtitles, and the potential for synchronisation with speech. However, when a pre-recorded segment follows a live section, the inevitable lag in the live subtitles means that the subtitling for the pre-recorded segments is also delayed. To synchronise the subtitling and speech for pre-recorded segments as soon as possible, broadcasters sometimes speed up the subtitling at the start. We would like to know whether subtitle users prefer this, or whether they find that speeded up subtitles are difficult to follow.

### **Editing of subtitles**

- 6.9 Truly verbatim subtitles are rare; almost all subtitles present an edited version of what is spoken. Natural human speech usually contains some redundant speech that can be dropped from subtitles. Sometimes speech can be paraphrased more economically, or less important details omitted, without losing the sense of what is being said. On occasions, the practical limits of how fast a subtitler can produce subtitles means that significant omissions have to be made, for example in chat shows, where people speak over others. You can get a flavour of this by looking at the figures in Section 5, which summarise the amount of editing by genres in the fourth set of samples. In some cases, it might be possible to provide more detail at the expense of longer lags in subtitling. We would like to know if subtitle users would prefer this, or not, and whether views differ among users with different levels of reading ability.

### **Maximum speeds of subtitling**

- 6.10 Some programmes contain a great deal of speech, some of it quite rapid (such as chat shows), while in others, there is less speech and the pace of delivery is more measured (such as entertainment programmes). The amount that can be subtitled

---

<sup>23</sup> Among others: Dhevi Rajendran, Andrew Duchowski, Pilar Orero, Juan Martinex and Pablo Romero-Fresco (2012), *Effects of text chunking on subtitling: A quantitative and qualitative examination*, in *Perspectives: Studies in Translatology*. Routledge Taylor & Francis Group (<http://andrewd.ces.clemson.edu/research/vislab/docs/perspectives.pdf>); Romero-Fresco, Pablo (2012): *Quality in Live Subtitling: The Reception of Respoken Subtitles in the UK* (<http://connection.ebscohost.com/c/articles/78120106/quality-live-subtitling-reception-respoken-subtitles-uk>); Romero-Fresco, Pablo (2011) *Subtitling through Speech Recognition*, Manchester: Routledge.

can be limited by two factors: first, how fast a subtitler can work; second, by rules on the maximum speed of subtitling, expressed in words per minute (wpm), or sometimes characters per second. Ofcom's current guidance (and some research) suggests that when speeds go beyond 180-200 wpm, subtitle users find the subtitles difficult to follow. But there are different views. We'd like to know what subtitle users think, and whether opinions vary between those with different levels of reading ability.

## Research

- 6.11 Ofcom intends to commission independent research on these issues, and any others that it identifies as significant in the light of comments from stakeholders. In the research we will aim to recruit subtitle users of different ages and reading ability, so that we can see whether there are any differences in how they view the trade-offs. As with other Ofcom-funded research, it will be carried out independently.
- 6.12 We have also begun to review third party research into issues that touch upon the experience of people using subtitles, so that we can take this into account.
- 6.13 Some issues, by their nature, do not readily lend themselves to qualitative research. For example, Action on Hearing Loss (AHL) has joined Ofcom in asking broadcasters to experiment with short delays in live programmes, so that this period could be used to produce synchronised 'live' subtitles, presented in blocks. While it is easy to imagine that this would benefit subtitle users, the negative trade-offs fall largely to the broadcasters, who argue that the cost and risks to them outweigh the benefits to the hearing-impaired. Qualitative research is not likely to shed much light on this, so Ofcom will continue with its efforts to persuade broadcasters to experiment with this proposition, so that more can be learnt about the actual drawbacks, and whether they can be managed and minimised.

## Review of Ofcom's Code and Guidance

- 6.14 We intend to discuss the results of the research with subtitle users and groups representing their interests, as well as broadcasters and access service providers during the first few months of 2016.
- 6.15 In the light of these discussions and lessons from third party research and our own research, we will look again at Ofcom's Guidance to broadcasters<sup>24</sup>, which has not been reviewed for ten years. We will also consider whether any changes to the Code on Television Access Services<sup>25</sup> may be appropriate – if made, these would be binding on broadcasters, as their licences (or in the case of the BBC, the BBC Agreement) require them to comply with the Code.
- 6.16 We would then invite views on any proposed changes, and take account of these in deciding whether to amend either document.
- 6.17 Finally, we will consider whether it would be helpful to carry out further monitoring exercises in future, whether in relation to the quality of live subtitling, or the number of pre-recorded programmes that have to be subtitled live, or the number of technical failures that have resulted in the loss or interruption of subtitling.

<sup>24</sup> Ofcom's Code on Television Access Services, 13 May 2015

(<http://stakeholders.ofcom.org.uk/binaries/broadcast/other-codes/tv-access-services-2015.pdf>)

<sup>25</sup> *ibid*, Annex 4.

## Annex 1

# External review of measurements



Pablo Romero-Fresco

Inma Pedregosa

1 October 2015

## Live Subtitling: A Qualitative Analysis of the Fourth Round of Measurements

### 1. Methodology

On 16th October 2013, Ofcom decided that broadcasters should be required to measure the following dimensions of quality, on the basis of samples of live subtitling selected by Ofcom:

- a) the average speed of the subtitling;
- b) the average latency of the subtitling (the delay between speech and live subtitling), and the range of latencies; and
- c) the number and type of errors (i.e. minor spelling errors, major omissions or factually misleading subtitles).

Ofcom identified samples of live subtitling in three genres of programming - news, entertainment and chat shows - and asked broadcasters to carry out measurements. Broadcasters were asked to collect data using the NER model, devised by Pablo Romero-Fresco (University of Roehampton) and currently used by regulators, broadcasters and subtitling companies in Australia, Spain, Germany, Switzerland, Italy and France, among other countries. In order to ensure consistency of measurements, a small team of experts at the University of Roehampton led by Pablo Romero-Fresco and Inma Pedregosa validated the measurements provided by broadcasters from a third-party standpoint, using the NER model. The results of the first, second and third round of measurements can be found on Ofcom's website<sup>26</sup>.

For this fourth round of measurements, the external reviewers have analysed a **total of 84 ten-minute clips** from 84 programmes belonging to three genres (news, entertainment and chat shows) and broadcast on five channels, namely BBC1 News at Six, The National Lottery: In It to Win It and The One Show from the BBC, Channel 4 News, Gogglebox and Sunday Brunch from Channel 4, Five News and Big Brother from Channel 5, ITV News, Britain's Got Talent and Loose Women from ITV and Sky News at Six, Game Changers and Soccer AM from Sky. In total, the analysis comprises 14 hours of live TV material including approximately 145,000 words and almost 23,000 subtitles.

---

<sup>26</sup> See: Ofcom, *The Quality of Live Subtitling* (<http://stakeholders.ofcom.org.uk/consultations/subtitling/>)

The results of the individual evaluations of every programme are not restricted to a single figure, in this case the accuracy rate. Instead, the NER model provides **a short assessment** of the quality of the subtitles for every programme, including the accuracy rate and also issues related to the delay of the subtitles, their position, their speed, their flow, the types of errors they contain, the way in which the speakers have been identified and the challenges posed by every programme, among other factors.

As far as the **accuracy rate** is concerned, the four rounds of assessments conducted so far confirm the figure of **98%** as a valid threshold from which subtitles may be considered of acceptable quality for TV broadcast. In our sample, and taking into account also the above-mentioned factors related to overall quality, subtitles with 99%-99.5% accuracy rate ranged from very good to excellent, whereas those with 98.5%-99% ranged from good to very good. Below those rates, the closer to 98% accuracy rate the subtitles were, the more problems they presented. For only two out of the 84 programmes analysed, the quality of programmes with accuracy rates below 98% was found to be acceptable on the basis of the other elements in the subtitles. In the rest of the programmes, the accuracy rate (and the threshold of 98%) was found to be in line with the overall quality assessment of the subtitles, including delay, position, speed, flow, speaker identification, etc.

The application of the NER model has proved very consistent and comparable across the internal reviewers from the different subtitling companies (who were only given a few written instructions as to how to apply the model) and the external reviewers from the University of Roehampton. The **average discrepancy** with regard to the accuracy rates of the 84 programmes from five different broadcasters is **0.07 percentage points**.

## 2. Results and discussion

On the whole, the **quality of the live subtitles** analysed in this report may be regarded as good, with an average accuracy rate of 98.55%, higher than the average accuracy rates obtained in the first (98.28%), second (98.34%) and third (98.37%) rounds of measurements. This fourth round has produced the most significant increase in average accuracy, confirming an upward trend that yields an overall accuracy rate of 98.38% across all four assessments.

Around 80% of the programmes analysed in this fourth round managed to reach the required quality threshold, as compared to 76% in the first analysis, 74% in the second (which was marred by an unusually high number of technical issues) and 77% in the third. In one third of the programmes included in this fourth round, the subtitles were good and in more than one fourth they were very good or excellent. In the present round, only 15% of the programmes presented technical issues, most of which involved subtitle freezes.

In general, the greater use or **pre-recorded subtitles** witnessed in the last two rounds of measurements for news and entertainment samples has contributed to better subtitling quality. The average accuracy rate of news subtitles has increased steadily from 98.49% in the first analysis to 98.62% in the second, 98.86% in the third and 99.02% in the fourth, in other words, from acceptable to very good quality. Their latency has decreased from an average of 5.7s in the first two rounds to an average of 4.6s in the last two rounds.

However, one of the issues highlighted in the previous round of assessments as a result of the use of this hybrid mode, the occurrence of excessive **peaks of speed**, is still widespread and is having a negative impact on the viewers' comprehension. The average speed of the subtitles in this sample is 133wpm, which should allow viewers more than enough time to read the subtitles and watch the images on the screen (approximately 40% of the time on the

subtitles and 60% of the time on the images)<sup>27</sup> and which falls below the highest recommended subtitling speed (180wpm-200wpm)<sup>28</sup>. However, in many news and entertainment programmes, the pre-recorded subtitles are often cued in too quickly in an attempt to catch up with the audio after a piece with delayed live subtitles. This issue, which affected 25% of the programmes in the second round and 99% of the programmes in the third round, has now been found in 92% of the programmes in this fourth analysis, including 92% of the chat shows, 83% of the entertainment programmes and all the news programmes in the sample. On average, chat shows have only 6 peaks of speed per programme (since they are mostly subtitled live without pre-recorded subtitles), whereas entertainment programmes have 14 peaks per programme and news programmes have 35 peaks per programme, roughly one third of the subtitles analysed in every programme. In 20% of the news programmes, at least half of the subtitles exceed 200wpm and news programmes in all channels have included subtitles over 300wpm. Although in some cases some subtitles have reached speeds over 350wpm, an effort has been made to keep these one-off peaks to a minimum, and thus the speeds over 500wpm found in the third sample have now been avoided.

At any rate, these fast subtitles cannot be processed by most viewers and normally correspond to the beginning of a news item as the programme goes back to the studio, which may also make it difficult for the viewers to understand the following subtitles for that particular item. We would recommend that the subtitling software is set to a maximum speed (ideally no higher than 180wpm-200wpm) for both pre-recorded and live scrolling subtitles. This will ensure that the subtitles can be read in full and that the subtitlers do not have to worry about measuring the time a subtitle must be on the screen, thus focusing on the other tasks they have to perform.

The increasing use of pre-recorded subtitles in live programmes also demands particular attention in cases in which there is **discrepancy** between what the presenters are saying and the information in the script that is used for the subtitles. This was found in twelve programmes in the sample analysed here (the previous round of assessment contained only four cases) and, depending on the seriousness of the error introduced in the subtitles, it may require a live correction by the subtitler, especially if it affects a figure that has been updated by the presenter.

In general, the combination of pre-recorded and live subtitles is very positive and has resulted in a significant increase in the overall quality of news and entertainment programmes. However, it is very important to solve these issues, especially the one related to speed, which has a serious impact on the viewers' comprehension and may end up defeating the purpose of introducing this hybrid mode.

As in the first three rounds of measurements, the overall subtitling quality of the programmes included in this sample varies greatly depending on the **genre** analysed, which is mainly due to the very different speech rates, content and the structure of the programmes.

The **news programmes** in the sample feature an average speech rate of 175wpm, as compared to 163wpm in the first round, 176wpm in the second and 184wpm in the third one. The content is more meaningful than that of entertainment programmes and chat shows, which makes these programmes particularly challenging to edit without losing important information. Moreover, some of the programmes included in this sample feature on-set interviews, which involve higher speech rates and overlapping dialogue, thus making the subtitlers' task considerably more difficult. Given the importance of content in this genre, subtitlers seem to be making a significant and commendable effort to avoid excessive editing. The average editing rate in this sample is 14%, very much in line with the average editing

---

<sup>27</sup> See Romero-Fresco, Pablo (2011) *Subtitling through Speech Recognition: Respeaking*, Manchester: St. Jerome.

<sup>28</sup> See Ofcom's "Subtitling –An Issue of Speed" (2005) and BBC's Subtitling Guidelines (2009).

rate of the first three samples (13.7%). This allows news subtitles to obtain a very good average accuracy rate of 99.02%, higher than in the first (98.49%), second (98.62%) and third (98.86%) samples. Whereas in the first three samples 33% of the news programmes analysed had very good subtitles and 10% did not meet the required quality threshold, in this last sample over 50% of the news programmes have very good subtitles and none of them falls below the threshold. As mentioned above, the only issues in these programmes are the above-mentioned peaks of speed, the discrepancies between the pre-recorded subtitles and the presenters' speech (in one in three programmes) and those cases where the subtitles obscure key information or the speakers' mouths on the screen (in 60% of the programmes).

As noted in the first two analyses, on average **entertainment programmes** normally have a lower speech rate than news and chat shows. They also combine live spontaneous dialogue with pre-recorded material and songs (thus providing much-needed breathers for subtitlers) and they lend themselves to a certain degree of editing without losing key content, since the spontaneous dialogue often features hesitations, false starts and redundant comments. The inclusion of Game Changers in the third sample and the addition of Big Brother in this fourth sample have altered this trend, since both programmes share many traits with a chat show. The average speech rate of the entertainment programmes in this sample (180wpm) is higher than in the first two samples (161wpm in and 156wpm, respectively). This has led to significant editing (an average of 30%) and to an average accuracy rate of 98.31%, lower than in the first (98.51%) and second (98.87%) analyses and very similar to the third analysis (98.34%). In order to improve the quality of the subtitles provided for these programmes, it is essential that broadcasters make available, whenever possible, the scripts for all pre-recorded material, which will result in higher accuracy rates and lower delays.

Once again, the **chat shows** are by far the most challenging programmes to subtitle live, not only because of the speakers' fast speech rates but also because of the presence of quick asides and overlapping speech that are essential to understand the humour in the programme. In the third sample, the average speech rate of the presenters and other people on screen was as high as 222wpm, thus triggering an editing rate of 41% and an average accuracy rate of 97.80%. In this fourth sample, the average speech rate is 185wpm. This has allowed for a more reasonable editing rate of 30% and an average accuracy rate of 98.26%, thus reaching for the first time the required threshold (in the first three analyses the accuracy rates for chat shows were 97.91%, 97.60% and 97.80%, respectively). Whereas in the third round of assessment, 40% of the chat shows did not reach the quality threshold, 50% had acceptable subtitles and 10% had good subtitles, in this last round, 33% of the chat shows do not reach the quality threshold, 25% have acceptable subtitles and 41% have good or very good subtitles. Be that as it may, the fact remains that one in three chat shows does not reach the required threshold, which means that despite the commendable effort made by the subtitlers, there seems to be no guarantee that the subtitles in these programmes can consistently provide good access to the viewers.

The average delay in this fourth sample is 5.6s, similar to the first and second samples (5.4s and 5.6s, respectively) and higher than the third one (5.1s). This increase is due to the above-mentioned inclusion of chat show-like entertainment programmes in the fourth sample (Game Changers and Big Brother), which has pushed up the average delay of the subtitles in entertainment programmes from 4.9s in the third sample to 5.9 in the fourth one. In contrast, the increasing use of pre-recorded subtitles in the news in the last two samples has enabled a significant reduction in the delay of the subtitles for these programmes, from 5.6s and 5.8s in the first and second assessments to 4.6s and 4.7s in the third and fourth.

As well as on the availability of pre-recorded subtitles, the delay of the subtitles also depends on the live subtitling technique used, the genre of the programmes and the occurrence of specific technical issues during transmission. When scripts are available and pre-prepared subtitles are combined with live subtitles, the delay has been reduced to 3-4s and for the first time some news programmes have average delays of 2-3s. In some cases, subtitles appear on screen slightly before or at the same time as the audio. The programmes or sections of programmes that only contain live subtitles produced by respeakers seem to have an average latency of 6s, with very good segments of 5s and even 4s but also peaks of 7s to 9s. In programmes or sections with many speakers, fast speech rates and overlapping speech, the average delay for these live subtitles is 7-8s, with good segments of 6s and peaks of up to 10s. Finally, regardless of the average delay of the subtitles and possibly due to specific technical problems, several programmes in the sample analysed feature peaks of 10-21-second delays and sometimes even higher, which should be avoided whenever possible.

As far as **errors** are concerned, the data obtained in this fourth analysis is very consistent with the first three. In the sample analysed, 69% of the errors included in the subtitles were editing errors, that is, those caused by incorrect omissions or additions made by the subtitlers, errors of speaker identification, spelling or punctuation mistakes, etc. The remaining 31% were recognition errors, those caused by the interaction between the subtitler and the steno machine or the speech recognition software. Once again, these percentages vary depending on the genres. In chat shows, which feature the highest speech rates, 74% of the errors are caused by incorrect editions and 26% by misrecognitions. Entertainment programmes, where speech rates are normally slower, contain 69% edition errors and 31% recognition errors. Finally, news programmes feature 65% edition errors and 35% recognition errors. This relative increase in recognition errors with regard to entertainment programmes and chat shows may be due to both the effort made by subtitlers to type/respeak fast in order to keep up with the audio without editing too much and to the very content of the news, which is likely to include more specialised terms or unexpected proper nouns than chat shows and entertainment programmes. This is one of the reasons why having access to the script before the programme – when one is available – can help improve accessibility for the viewers.

As far as the **seriousness of the errors** is concerned, in general 56% of the errors found in the sample are minor (i.e. they do not prevent the viewers from following the content of the programme), 39% are standard (they trigger confusion or cause full factual omissions) and 5% are serious (they introduce misleading information). These figures vary depending on whether they relate to edition errors (54% of which are minor, 42% are standard and 4% are serious) or recognition errors (61% of which are minor, 29% are standard and 10% are serious). In other words, not only are edition errors more numerous than recognition errors but they also tend to be more problematic, since it is more common to have standard edition errors (omissions of full sentences) than standard recognition errors (nonsensical misrecognitions). The different genres also play an important role here. As a result of the effort made by the subtitlers to improve the quality of the subtitles for the news, one third of the errors in these programmes is standard and two thirds are minor. In contrast, in chat shows (and in this sample in entertainments programmes too, especially due to the inclusion of Game Changers and Big Brother) the fast speech rates and the overlapping interventions of the speakers force the subtitlers to rush and to omit more information. As a result, 53% of the edition errors are minor and as many as 42% are standard. In other words, almost one in two errors found in chat shows and entertainment programmes involves the omission of a full sentence and may cause the viewers to lose the thread of the programme.

Other issues identified in this fourth analysis are:

- the presence of **subtitles that obscure** key information on the screen or the speakers' mouths, thus making it impossible for viewers to lip read. This was found in one in five programmes in the third sample and has increased to 50% of the programmes in the present sample. In contrast, some subtitlers have managed to place the subtitles on top or to change their position live so as not to obscure important information.
- the **omission of swearwords** in the subtitles of two programmes;
- the use of the **wrong colour** for character identification in a few programmes;
- a few cases of **bad segmentation**;
- a few excessively **late corrections**.

### 3. Conclusions of the last assessment

The quality of the live subtitles analysed in this report is good, with the highest accuracy rate obtained in all four rounds of assessment (98.55%). On the whole, 80% of the programmes analysed in this fourth round managed to reach the required quality threshold. In one third of the programmes, the subtitles were good and in more than one fourth they were very good or excellent. This may be explained by the fact that there were fewer technical issues than in the second round and especially by the increased use of the hybrid mode, i.e. the combination of live and pre-recorded subtitles for news and entertainment programmes. The most noticeable increase in quality can be seen in the subtitles for news programmes, whose accuracy has increased from 98.5% to more than 99% over the four rounds of assessment and whose average latency has decreased from 5.6s to 4.7s.

However, one of the issues highlighted in the previous round of assessments as a result of the use of this hybrid mode, the occurrence of excessive peaks of speed in the transition between live and pre-recorded subtitles, is still widespread and is likely to have a negative impact on the viewers' comprehension. In the present sample, 92% of all the programmes feature speeds over the recommended 200wpm, including 92% of the chat shows (which have an average of only 6 peaks per programme), 83% of the entertainment programmes (with an average of 14 peaks per programme) and all the news programmes (with an average of 35 peaks per programme). We would recommend that the subtitling software is set to a maximum speed (ideally no higher than 180wpm-200wpm) for both pre-recorded and live scrolling subtitles. This will ensure that the subtitles can be read in full and that the subtitlers do not have to worry about measuring the time a subtitle must be on the screen, thus focusing on the other tasks they have to perform.

For the first time since this project began, the subtitles produced for chat shows have managed to reach the required threshold, with an average accuracy rate of 98.26%. Around 25% of the chat shows in the sample have acceptable subtitles and 41% have good or very good subtitles. However, the fact remains that the subtitles in one in three chat shows do not reach the required threshold. Despite the subtitlers' best efforts, the characteristics of these shows and the absence of a script make it very difficult to guarantee consistently good subtitles for these programmes. The same goes for some pre-recorded entertainment programmes such as Game Changers and Big Brother (which share many traits with chat shows) and for Gogglebox, whose constant switching between scenarios and speakers, added to the unavoidable 5- or 6-second delay of the subtitles, makes it almost impossible for the viewers to identify who is speaking, no matter how good the subtitles are.



To conclude, this fourth round of measurements has confirmed the findings obtained in the third round, that is, the need to control the subtitling speed in the new hybrid mode and to make available, whenever possible, the scripts for chat shows, entertainment and news programmes. Last but not least, it has shown further evidence of great work by the subtitlers, who, despite the complexity of their task, have managed to improve the live subtitles they produce and to find solutions for the new issues that have arisen throughout these two years.