



Measuring live subtitling quality

Results from the second sampling exercise

Publication date: 5 November 2014

Contents

Section		Page
1	Summary	2
2	Accuracy	7
3	Latency	10
4	Speed of subtitling	13
5	Edition rates	16
6	Related issues	19
Annex		Page
1	External review of measurements	21
2	Additional Charts	27

About this document

This document is the second of four reports on the quality of live subtitling in British television programmes, based on samples drawn from live-subtitled programming broadcast in April and May 2014 by the BBC, ITV, Channel 4, Channel 5 and Sky.

In order to address continuing complaints about the quality of live subtitling, Ofcom consulted in May 2013 on proposals to incentivise broadcasters and access service providers to identify and act upon areas for improvement. Following this consultation, Ofcom published a statement in October 2013 explaining that we would ask broadcasters to measure and report on the key dimensions of the quality of live subtitling – accuracy, latency and speed – as well as on other related issues.

This report contains an update on the progress made in by broadcasters in other areas that affect the quality of live subtitling, such the use of block subtitling. The next report is scheduled for April 2015.

Section 1

Summary

Introduction

- 1.1 This document is the second of four reports on the quality of live subtitling in British television programmes, based on samples drawn from live-subtitled programming broadcast in April and May 2014 by the BBC, ITV, Channel 4, Channel 5 and Sky.

Background

- 1.2 In the light of continuing concerns about the quality of live subtitling and our statutory duties, Ofcom consulted in May 2013¹ on proposals to incentivise broadcasters and access service providers to identify and act upon areas for improvement. Following this consultation, Ofcom began a project² to measure the quality of live subtitling, in order to identify areas for improvement and encourage broadcasters to act upon these.
- 1.3 The first report was published in April 2014³. On the basis of samples of live subtitling selected by Ofcom, broadcasters were asked to measure the following dimensions of quality:
- a) accuracy: the number and type of errors (i.e. minor spelling errors, major omissions or factually misleading subtitles), using the NER model⁴;
 - 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.
- 1.4 These measurements were then checked for consistency by external reviewers from the University of Roehampton.

Second report

- 1.5 For the second report, we asked broadcasters to measure the quality of samples from live programmes broadcast in April and May 2014. We report on the accuracy, latency and speed of subtitling in these samples in the remainder of this document. In addition, we report on issues raised by the reviewers, on the greater use made by broadcasters of easier-to-use block subtitles, and on the extent to which subtitles are edited down.
- 1.6 Sections 2, 3 and 4 of this document report on the results for accuracy, latency and speed respectively, as well as specific points made by the external reviewers. The

¹ Ofcom, *The quality of live subtitling – improving the viewer experience*, 15 May 2013 (<http://stakeholders.ofcom.org.uk/consultations/subtitling>)

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

³ *Measuring live subtitling quality – Results from the first sampling exercise*, Ofcom, 30 April 2014 (<http://stakeholders.ofcom.org.uk/binaries/consultations/subtitling/statement/sampling-report.pdf>)

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

data was also used to assess the extent to which subtitlers edited down speech, and the results are summarised in Section 5. Finally, Section 6 reports on the progress made by broadcasters in making greater use of easier-to-read block subtitling.

Key points

- 1.7 Co-operation between broadcasters and access service providers appears to be improving. Broadcasters have reported making greater use of pre-prepared block subtitling in live programmes. The external reviewers estimate that a significantly higher proportion of live programmes in the second sample contained pre-recorded subtitling than in the first sample. This is easier to read, and is likely to be more accurate. The ability to pre-prepare block subtitles does, of course, depend on access to running orders, scripts and video packages. Ofcom welcomes the progress in this area.
- 1.8 The commonly-used threshold for ‘acceptable’ accuracy is 98%. In the second set of samples, **accuracy** was generally good, but rather variable. Accuracy rates by genre and broadcaster are shown in the graphs in section 2. Key points to note:
- the use of pre-prepared subtitling for pre-recorded elements of live programmes contributed to very good levels of accuracy in the samples drawn from news and entertainment programming, which had median accuracy rates of 98.8% and 98.9% respectively;
 - there were more samples in the second round that showed very good to excellent accuracy, defined as an accuracy rate of 99% or higher. However, the proportion of samples that failed to meet the 98% quality threshold rose from 20% to 25%;
 - more than half of the samples for chat shows failed to reach 98%, which means in practice that they would be difficult for subtitle users to follow;
 - greater use of pre-prepared subtitling allowed more of it to be presented in block subtitles, which research has shown are easier to read, and which allow viewers more time to view other content;
 - in some cases, the benefits from an increase in accuracy were lost, as pre-prepared subtitling was cued out too quickly to be intelligible, as explained by the external reviewers in Annex 1. Ofcom believes that these teething problems can be addressed, and looks to broadcasters to do so; and
 - in some cases, significant technical problems meant that the subtitling was unavailable for portions of the programmes sampled, rendering it of unacceptable quality.
- 1.9 **Latency** continues to be problematic; this round’s samples were generally worse than the maximum recommended 3 seconds in Ofcom’s guidelines. Measurements by genre and broadcaster are shown in Section 3. At 5.8 seconds, the median latency is 0.2 seconds longer than the median latency for the first set of samples. There were spikes of up to 21 seconds in some cases, and only three samples had a median latency of less than 3 seconds.
- 1.10 The use of pre-recorded subtitles helped to reduce delays in those segments where scripts were available (see Section 6); nevertheless, when subtitling had to be done live, the average latency was significantly higher (about 6 seconds).

- 1.11 The median **speed** measured in this exercise (146wpm) was well below Ofcom's maximum recommended range of 160-180 words per minute for pre-recorded subtitling (Section 4). However, segments combining pre-recorded and live subtitles were often characterised by peaks in subtitling speed. As the external reviewers explain in Annex 1, this appeared to stem from moves to synchronise pre-recorded subtitles with the audio.
- 1.12 In particular, the external reviewers noted that, at the junction between live subtitles and pre-recorded subtitles, operators have cued the pre-recorded subtitles too quickly, in an attempt to catch up with the audio. This caused peaks in subtitling speed and technical difficulties which affected the quality and readability of the live subtitling. In those instances, occurring in a quarter of news and entertainment samples, the subtitling speed went far beyond Ofcom's recommended range, with recorded speeds as high as 460wpm, which is practically unreadable.
- 1.13 Ofcom understands that there is readily available software that can be used to limit the speed at which pre-prepared subtitling is cued out, so that it remains intelligible, while reducing the intervals between subtitling in order to move smoothly towards synchronisation. Ofcom encourages broadcasters to address these teething problems, so that viewers can enjoy the benefits of both more accurate subtitling, and reduced latency.

Related issues

Technical problems

- 1.14 Technical problems bedevilled a number of programmes in the second set of samples.
- 1.15 Channel 4's *Sunday Brunch* was the worst affected programme:
- a) on 6 April, subtitling disappeared on two occasions for 1m 20s at 12:17:05 and 12:20:43;
 - b) on 13 April, there were many omissions, as well as subtitles disappearing too quickly or freezing, and latency peaking at 10 seconds (over three times Ofcom's recommended limit);
 - c) on 4 May, there were similar problems, with subtitles freezing for up to 25 seconds on one occasion, and old subtitles being retained while new ones were added;
 - d) on 25 May, subtitling was lost on three occasions, on each occasion for more than a minute.
- 1.16 There were also interruptions to subtitling on Channel 4's *Googlebox* on 18 April, and *Channel 4 News* on 15 May. Channel 4 informed us that its access services provider is currently conducting an investigation to identify the source of the problem. However, Channel 4 explained that subtitles are distributed separately for every platform; therefore, until the source of the glitches is identified, it is not possible to know whether the technical problems affected the subtitling shown on all platforms, or whether this issue appeared only on the platform used to record the samples.

- 1.17 Finally, BBC1's *The One Show* on 12 May experienced an interruption of the subtitles for 40 seconds at 18:38; the BBC explained that this was due to a technical issue with the subtitling software.
- 1.18 We have asked broadcasters to report on the number of technical failures resulting in the loss of subtitling during the period November 2014 to January 2015, and we shall publish the results in spring 2015.

Greater use of block subtitling in live programmes

- 1.19 At the beginning of this exercise⁵, we encouraged broadcasters to increase the use of the block subtitles on live programming, given the research suggesting that subtitle users find this format much easier to read than scrolling subtitles.
- 1.20 For the purposes of the second report, we asked broadcasters to provide us with an update on their use of block subtitles on live-subtitled programming. Individual broadcasters' responses are summarised in Section 6. All broadcasters have reported to us that the provision of block subtitling in live programming has increased substantially over the past six months, which is a most welcome development.
- 1.21 This has been achieved through greater cooperation between broadcasters, producers and subtitle providers, mainly by making material such as running orders or scripts available to subtitlers in advance. Ofcom understands that there remain cases where this is not yet happening, and encourages broadcasters to secure the necessary co-operation.
- 1.22 As explained above, greater use of block subtitling has given rise to teething problems that will need to be addressed in order that improvements in presentation are not secured at the expense of speeds that make subtitling practically unreadable. Ofcom is confident that broadcasters and their access service providers will do so.

Short delays to improve subtitling quality

- 1.23 In our consultation in May 2013, we invited consultees to comment on the potential of inserting short delays in live transmissions sufficient to improve the quality of subtitling⁶. Such delays are routinely used by public service broadcasters in the Netherlands and Flanders to eliminate latency and improve accuracy.
- 1.24 Our statement in October 2013 reported that each of the broadcasters – the BBC, ITV, Channel 4, Channel 5 and BSkyB - were strongly opposed to such a practice, citing risks to viewer trust, the need for complex technical solutions, and competition from other media, amongst other reasons⁷.
- 1.25 We decided nonetheless to discuss this issue further with broadcasters, as it is not obvious how else broadcasters will be able to tackle latency, which is a top of mind issue with viewers. We asked some of the broadcasters if they would experiment with inserting delays, initially in pre-recorded programmes that would have to be subtitled live, either because they were delivered late, or because they were highly topical.

⁵ Ofcom, *Measuring the quality of live subtitling: Statement*, 16 October 2013; paragraphs 3.96-3.99.

⁶ Ofcom, *The quality of live subtitling – improving the viewer experience*, 15 May 2013; see paragraphs 6.24-6.45.

⁷ Ofcom, *Measuring the quality of live subtitling: Statement*, 16 October 2013; paragraphs 3.104-3.106

- 1.26 In response, broadcasters reiterated their opposition to experimenting with short delays in transmission to improve the quality of subtitling, citing the same reasons as previously⁸. We shall be asking for further information to help evaluate the concerns broadcasters have expressed, and will provide feedback on this in a subsequent report.

Next steps

- 1.27 The next sampling exercise will take place in November 2014, drawing on news, chat show and entertainment programmes broadcast in October and November 2014 with live subtitles. We plan to report on the outcome in spring 2015. This report will include updates on the number of pre-recorded programmes that were delivered late and so had to be subtitled live during the period July to November 2014, the number of technical failures resulting in the loss of subtitling during the period November 2014 to January 2015, and an update on the issue of inserting short delays in live-subtitled programmes.
- 1.28 The final set of samples will be taken in May 2015. We expect to report on the results in Autumn 2015. This final report will pave the way for consideration of any future action that may be indicated, including a review of Ofcom's guidance.
- 1.29 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.

⁸ Ofcom, *Measuring live subtitling quality – Results from the first sampling exercise*, 30 April 2014; paragraphs 6.11-6.15.

Section 2

Accuracy

- 2.1 Accuracy is commonly identified as the key element of the quality of live subtitling by hearing-impaired viewers. While some mistakes are inevitable, due to the complexity of the respeakers' task and to technical issues, they force viewers to spend time working out what was meant. In some situations, even minor errors such as the substitution of 'fifteen' for 'fifty' may lead to misinformation or confusion for the viewer.
- 2.2 The team of external reviewers identified 98% as the quality threshold above which the quality of subtitling can be considered as 'acceptable'.

Summary of results

- 2.3 Figures 1, 2 and 3 below show median accuracy rates for the first and second sets of samples taken from news bulletins, chat shows and entertainment programmes. For reference purposes, the graphs below show the 98% acceptable quality threshold with a bold red line. Charts showing comparisons of minimum and maximum accuracy scores between the first and second rounds of measurements are available in Annex 2.
- 2.4 In general, median measurements show that accuracy rates were acceptable or better for both news and entertainment programming, but remained poor for chat shows. In two of the ITV programmes analysed, although overall accuracy rates were above 98%, some of the subtitles were rendered unintelligible due to technical problems.
- 2.5 The external reviewers noted that the proportion of serious edition errors⁹ has decreased, representing only 1% of all edition errors. In their view, this evidences a greater effort on the part of subtitlers to keep up with the live audio and omit, whenever required, only minor parts of the speech as opposed to full sentences.

⁹ These are the result of the respeakers' judgement or decision when they edit the original speech in order to reduce the total number of words or the complexity of the subtitles, as opposed to recognition error which are caused by the subtitling software.

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

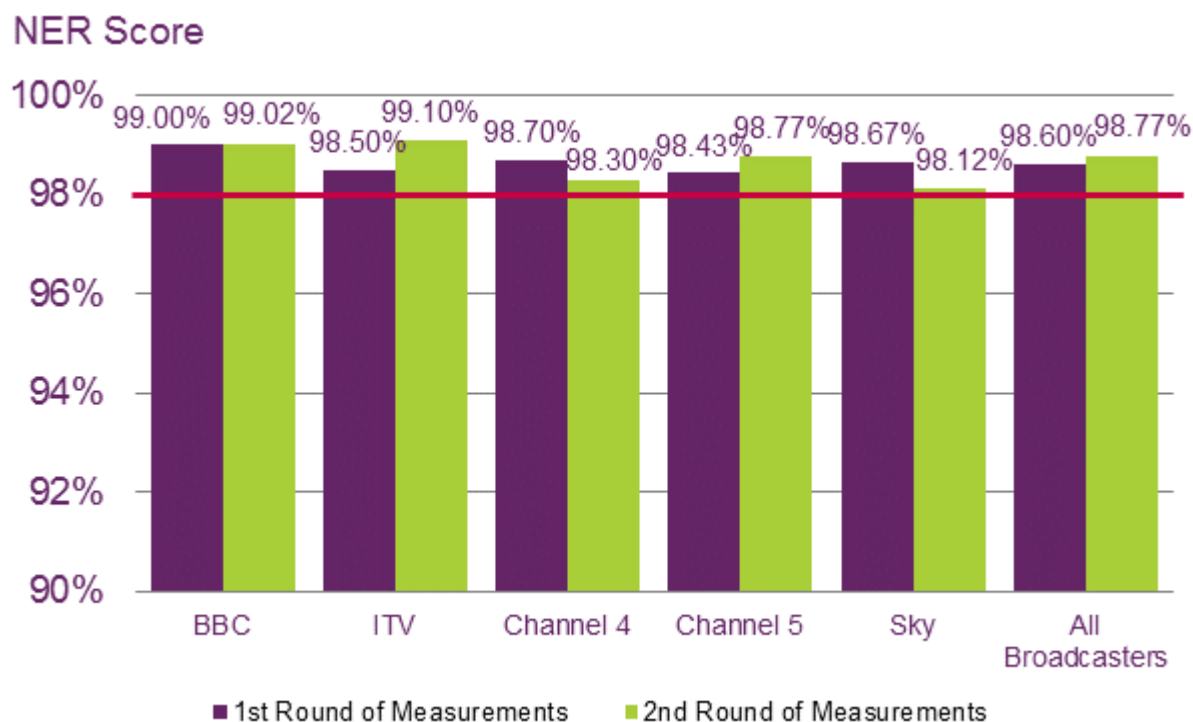


Figure 2: median accuracy rates in samples drawn from chat shows
 (higher bars represent a better viewing experience)

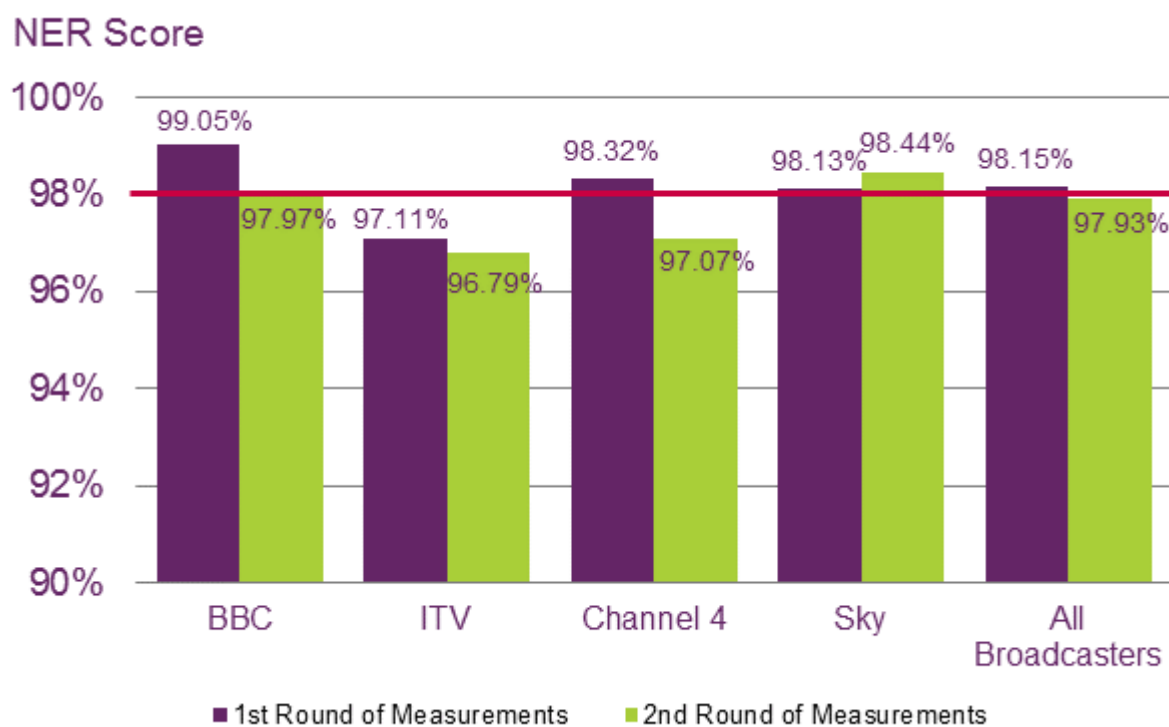
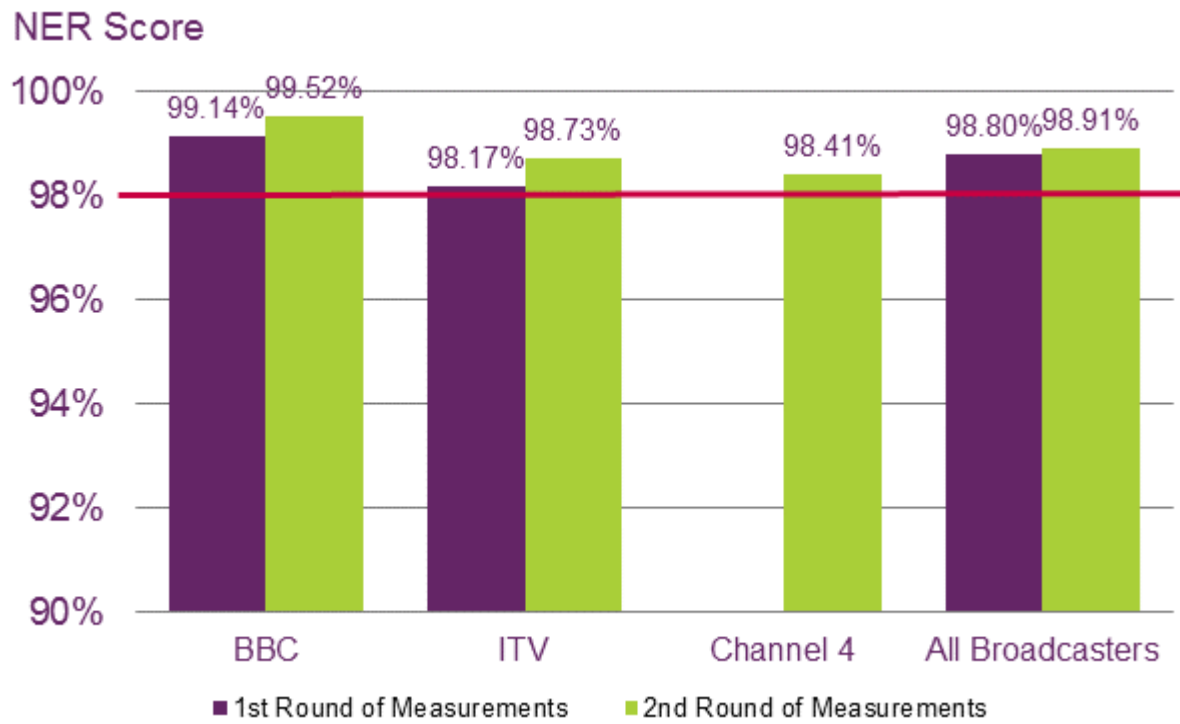


Figure 3: median accuracy rates in samples drawn from entertainment programmes
 (higher bars represent a better viewing experience)



(the first round of measurements did not include samples of entertainment programming shown by Channel 4)

Section 3

Latency

3.1 In the context of this project, we use the term ‘latency’ to indicate the delay between audio and subtitling. Ofcom’s guidelines¹⁰ recommend that the delay in subtitling presentation on live programmes should be no greater than three seconds. Ofcom will review its guidance in the light of the data gathered from the current exercise.

Summary of results

- 3.2 Figures 4, 5 and 6 below show median latency in the samples taken from news bulletins, chat shows and entertainment programmes. For reference purposes, the recommended maximum 3 second delay is marked on the graphs below with a bold red line. Charts showing comparisons of minimum and maximum latency measurements between the first and second rounds of measurements are available in Annex 2.
- 3.3 The external reviewers noted overall average latency was slightly higher than the first round of measurements, up from 5.5 to 5.7 seconds. This was despite the much wider availability of advance scripts and other material used to create pre-recorded subtitles. Only three samples out of 72 had a latency lower than the 3 seconds recommended by Ofcom’s guidelines.
- 3.4 The external reviewers noted that combining pre-recorded and live subtitles helped to reduce latency on segments where scripts were provided to subtitlers; however, the live subtitles were delayed for longer, with peaks of 7 to 9 seconds. This often resulted from high speech rates and overlapping speeches.
- 3.5 Finally, the external reviewers identified several instances in which, possibly due to technical problems, the subtitles appeared with a delay longer than 10 seconds, with peaks as high as 21 seconds.

¹⁰ Ofcom’s Code on Television Access Services, 18 October 2012; Annex 4, paragraph A4.18 (<http://stakeholders.ofcom.org.uk/binaries/broadcast/other-codes/tv-access-services-2013.pdf>).

Figure 4: median latency in samples drawn from news bulletins
 (higher bars represent an inferior viewing experience)

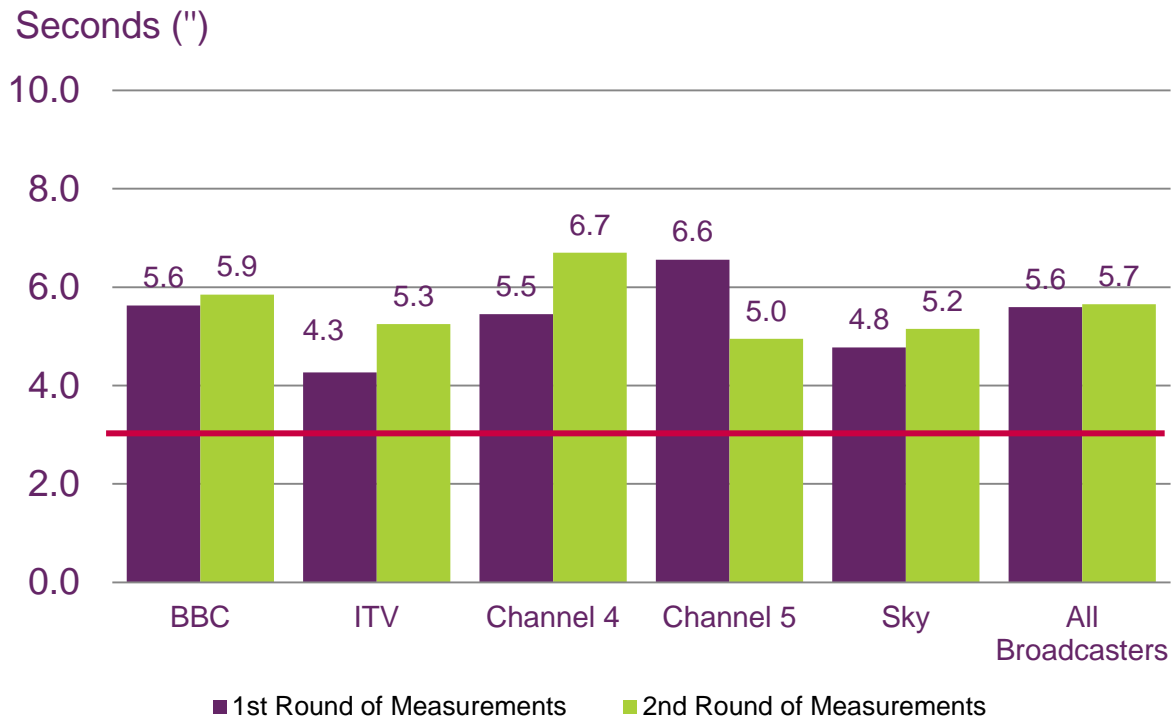


Figure 5: median latency in samples drawn from chat shows
 (higher bars represent an inferior viewing experience)

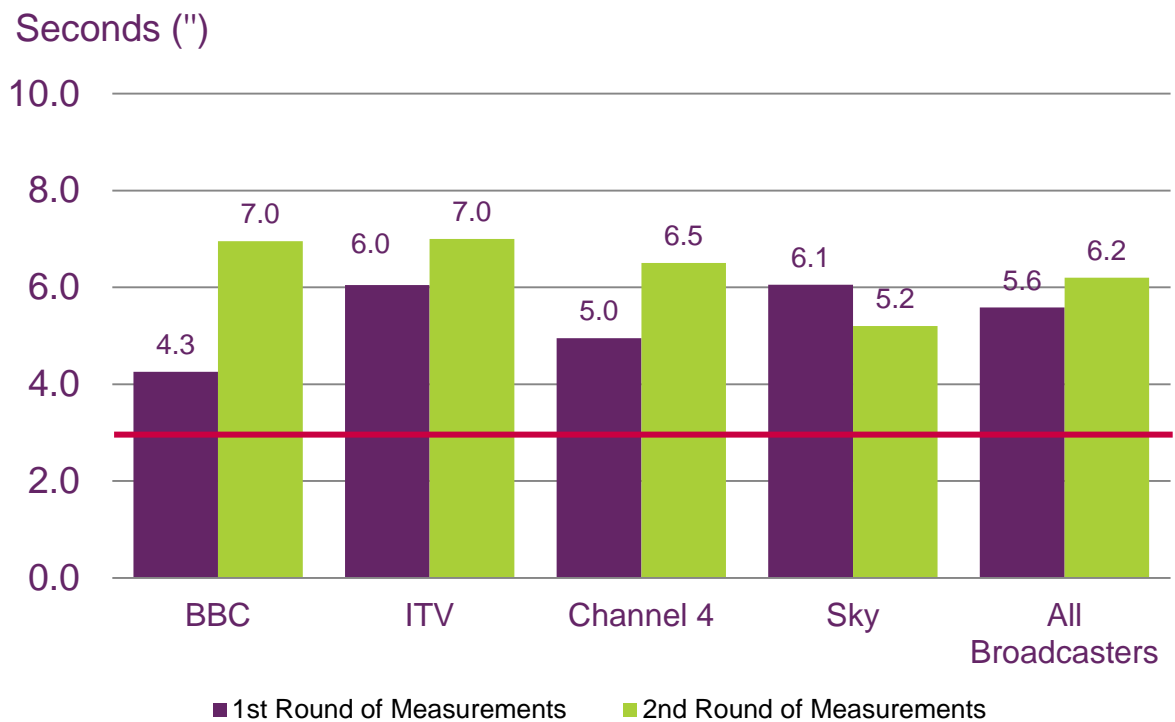
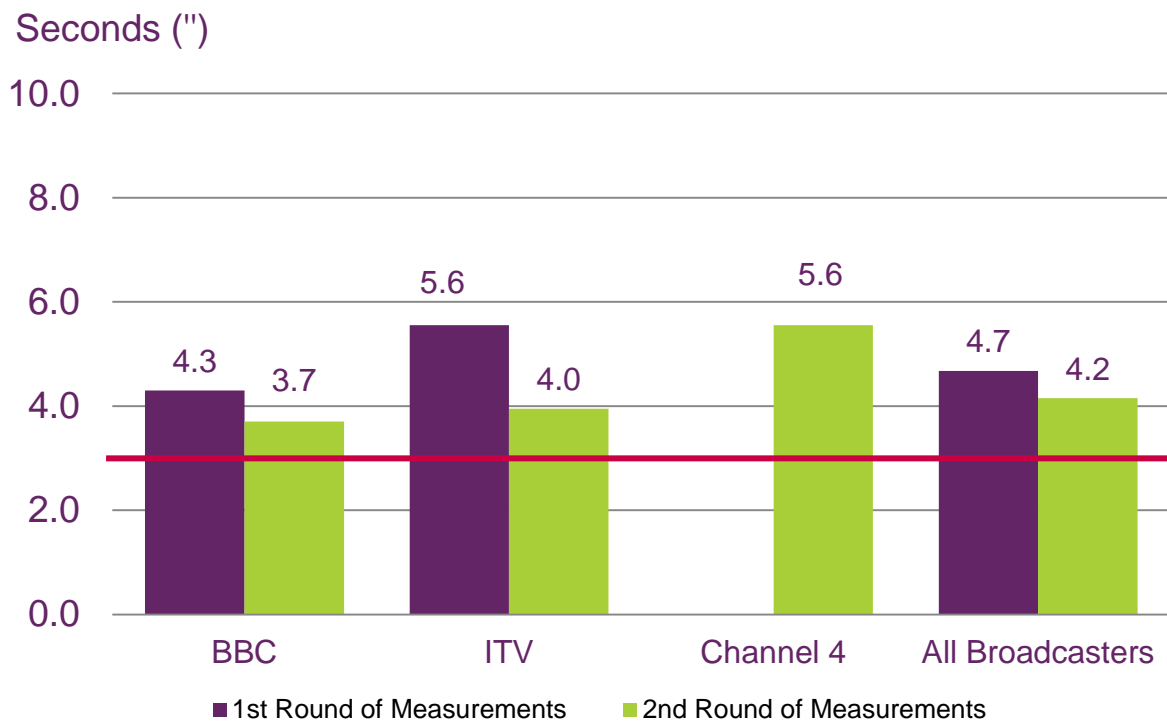


Figure 6: median latency in samples drawn from entertainment programmes
(higher bars represent an inferior viewing experience)



(the first round of measurements did not include samples of entertainment programming shown by Channel 4)

Section 4

Speed of subtitling

4.1 For the purposes of this project, the speed of subtitling is measured in words-per-minute ('wpm') rates. Ofcom's guidelines recommend that 'the speed should not normally exceed 160 to 180 wpm for pre-recorded programmes'¹¹. In the past, it was considered impracticable to restrict the speed of subtitles for all live programmes. However, Ofcom understands that commercially available software can be used to limit the speed at which pre-prepared subtitling is cued out, so that it remains intelligible, while reducing the intervals between subtitling in order to advance towards synchronisation.

Summary of the results

4.2 Figures 7, 8 and 9 below show the median speed of subtitling in the samples taken from news bulletins, chat shows and entertainment programmes. For reference purposes, the upper and lower boundaries of the maximum range recommended for pre-recorded subtitling is indicated on the graphs below with bold coloured lines. Charts showing comparisons of minimum and maximum latency measurements between the first and second rounds of measurements are available in Annex 2.

4.3 As in the first round of measurements, the speed of most samples of live subtitling was below the recommended maximum band of 160-180 wpm. Only in 13 instances was the subtitling speed higher than 160 wpm, though no sample showed an average speed rate above 180wpm. Overall, speed was highest on news bulletins than in other genres, due to the highest speech rate which is used to condense as much information as possible in relatively short periods of time, and the attempt by subtitlers to edit as little as possible.

4.4 Although average subtitling speeds are not a cause for concern, they conceal excessive speeds in a number of news and entertainment programmes. In most cases, this appeared at the junction between scrolling subtitles (always delayed) and pre-prepared subtitles (potentially synchronous), where the subtitler attempted to synchronise the pre-recorded subtitles to the audio as quickly as possible, when these followed a segment with naturally delayed live subtitles. Very high peaks in subtitling speed were found in around a quarter of all news and entertainment programmes, with speeds of 290, 350 and even 460 wpm, which are unreadable for most viewers.

4.5 The combination of live and pre-prepared subtitles, and the potential problems to which these give rise, are not unique to the UK. Ofcom is aware that some European broadcasters manage the transition by using automated software to cue out block subtitles, with hard limits on the speed, and the ability to shorten the gap between blocks of subtitles, in order to reduce or eliminate latency in a controlled fashion that keeps subtitles at a readable speed. Ofcom understands that some (if not all) access service providers have this capability.

4.6 Furthermore, it is our understanding that the current technology allows pre-recorded subtitles to be cued out automatically, at pre-fixed maximum speeds. This would reduce the tendency, witnessed in this sampling exercise, to accelerate the transmission of pre-recorded subtitles in order to synchronise them with the original

¹¹ *Ofcom's Code on Television Access Services*, 18 October 2012; Annex 4, paragraph A4.19

speech. Ofcom hopes that this technology could be implemented to reduce the occurrence of peaks in subtitling speed.

- 4.7 Finally, this sampling exercise has highlighted the risk of aggregate speed measures – such as the one used for this report – missing some granular information about subtitling speed and the impact of peaks on the quality of live subtitling. On the basis of the information we asked broadcasters to provide for the purposes of this exercise, external reviewers were only able to identify a portion of the total peaks in speed which make subtitles very hard if not impossible to read.
- 4.8 For this reason, from the next sampling exercise, we shall be asking broadcasters to provide us with time-coded transcripts of both the original audio and the subtitles. Only in this way will it be possible for external reviewers to identify and report on all the instances in which the subtitling speed went well beyond the recommended guidelines.

Figure 7: median speed of subtitling in samples drawn from news bulletins

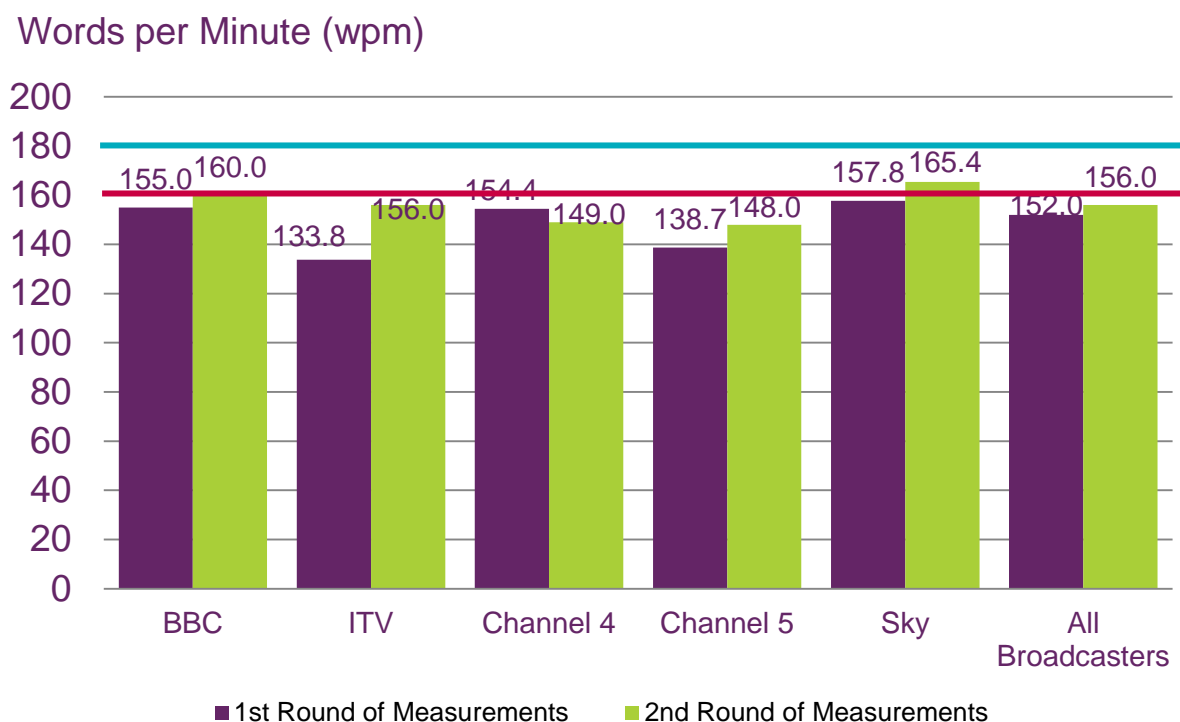


Figure 8: median speed of subtitling in samples drawn from chat shows

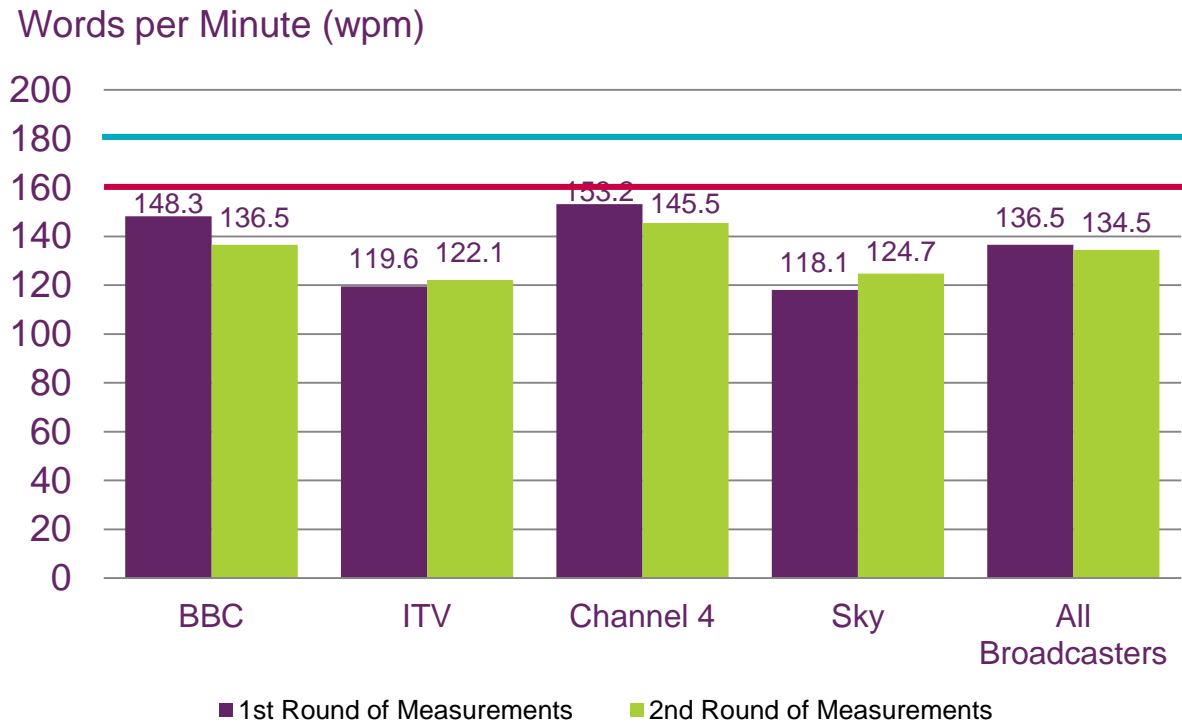
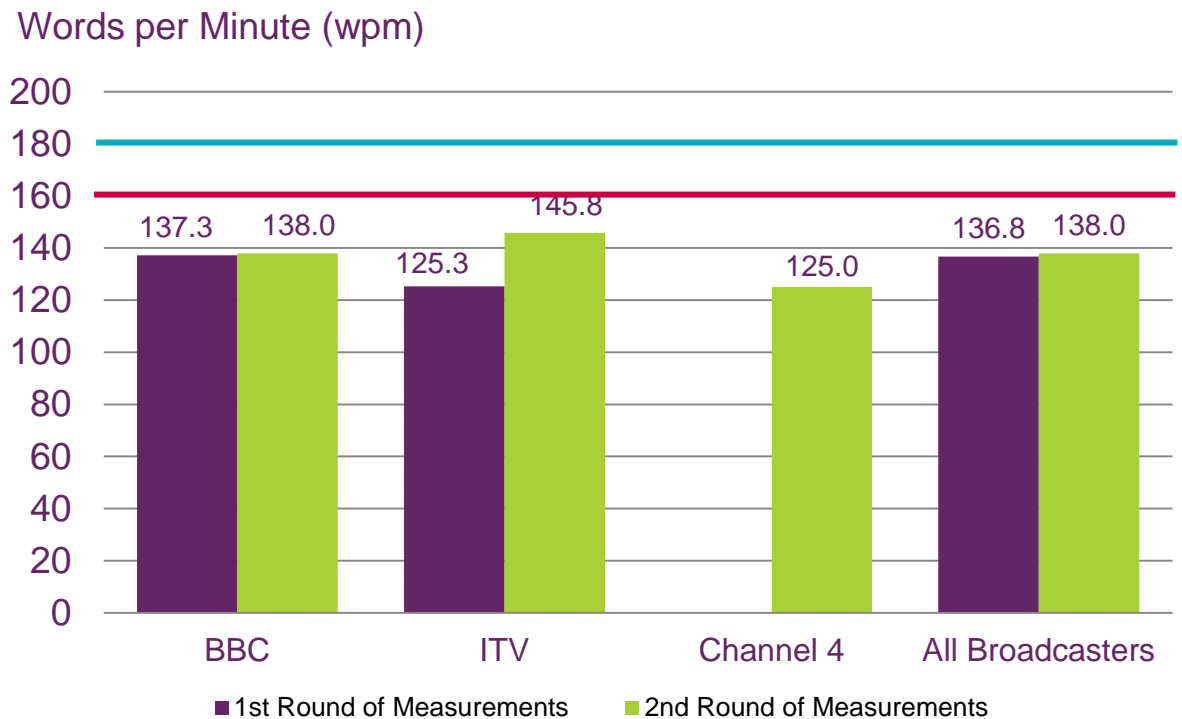


Figure 9: median speed of subtitling in samples drawn from entertainment programmes



(the first round of measurements did not include samples of entertainment programming shown by Channel 4)

Section 5

Edition rates

5.1 The edition rate provides a measure of the extent to which the subtitles in the sample represent an edited version of what was said. Given that the edition rate is not a measure of quality, Ofcom has not issued any specific guidance on what the appropriate edition rate should be. 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, and those who advocate greater amounts of editing in order to keep the speed of subtitling down.

Summary of results

5.2 Figures 10, 11 and 12 summarise the median edition rates by genre and broadcaster. The median edition rate in the whole sample was 16.8%, only slightly higher than the median edition rate measured in the first sampling exercise (15.14%).

5.3 It is important to note that edition rates differ significantly across genres; in that respect, the external reviewers¹² noted that:

- a) “the news programmes in the sample ...normally have more meaningful content [than entertainment programmes], which is particularly challenging to edit without losing important information ... Given the importance of content in this genre, subtitlers seem to be making a significant and commendable effort to edit less than in entertainment programmes (an average editing rate of 13.3)”;
- b) “chat shows are the most challenging genre to tackle as far as live subtitling is concerned, not only because of the speakers’ fast speech rates ..., but also because of the presence of overlapping speech that is essential to understand the humour in the programme. The subtitlers have thus been forced to edit an average of 35% of the original content”; and
- c) “entertainment programmes ... 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”. Furthermore, “the combination of pre-recorded and live subtitles enables the subtitlers to keep more information from the audio in the subtitles (the average editing rate is 15.77%, compared to 20% in the previous analysis)”.

¹² N.B. External reviewers use averages rather than median to indicate the typical edition rate in the sample; see Annex 1 for the external reviewers’ entire report.

Figure 10: median edition rate in samples drawn from news bulletins

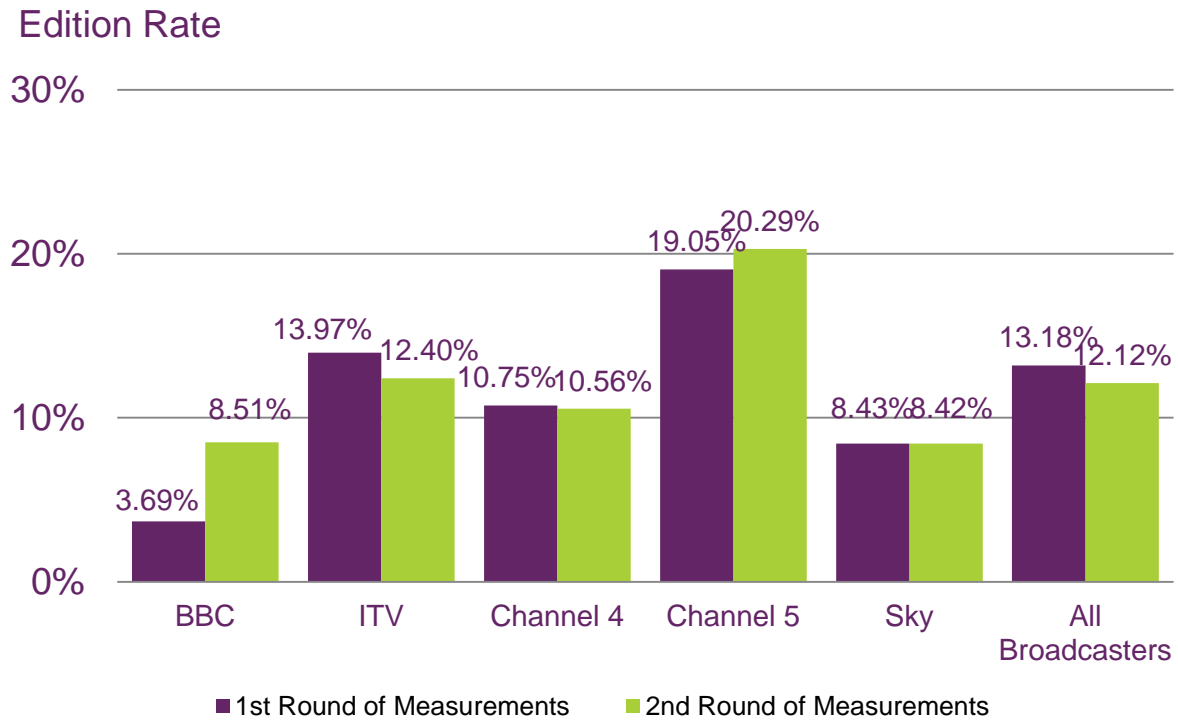


Figure 11: median edition rate in samples drawn from chat shows

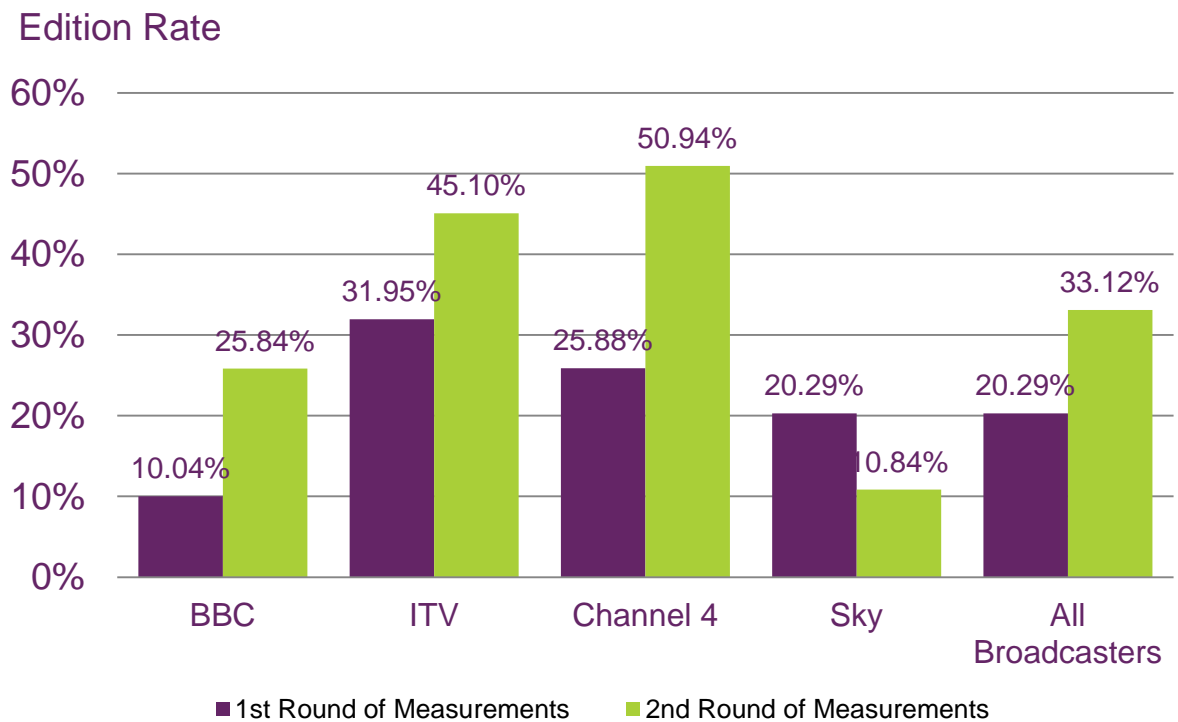
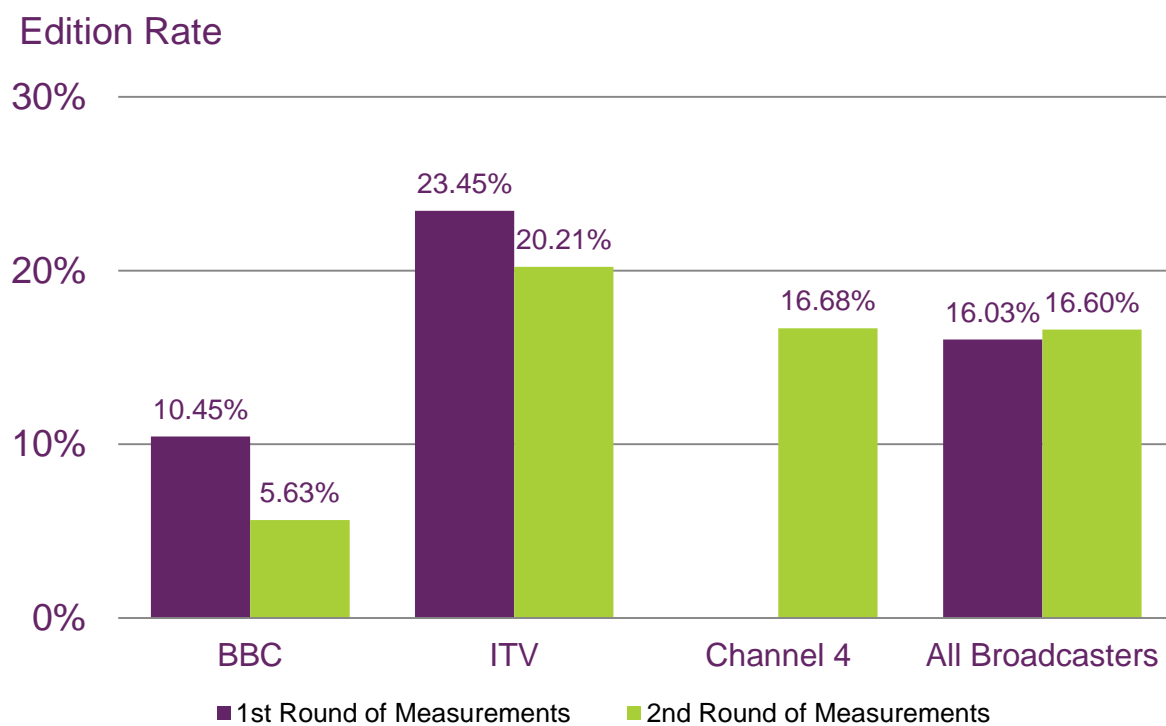


Figure 12: median edition rate in samples drawn from entertainment programmes



(the first round of measurements did not include samples of entertainment programming shown by Channel 4)

Section 6

Related issues

Greater use of block subtitles in live programmes

- 6.1 In our statement of October 2013, we encouraged broadcasters to increase the use of block subtitles, given evidence suggesting that these are much easier for viewers to read and comprehend. Broadcasters have committed to using block subtitles whenever possible, and indicated how this practice would become easier with the upcoming upgrades of subtitling software.
- 6.2 For the purposes of this report, we have asked broadcasters for an update on their use of block subtitling over the past six months:
- a) the BBC said that it had been working closely with their access service provider to provide block subtitles on more live content; since September 2014, the BBC had made newsroom running orders, studio scripts and news packages prepared in advance available to the subtitling provider ahead of transmission for all live programming where this material is available. The BBC also noted, however, that whenever this material is not available, it will continue to use the scrolling format of subtitles.

The BBC also said that it is increasing the reversioning of live scrolling subtitles into block subtitles in programmes that are repeated.

- b) ITV said that in the past six months there have been significant improvements in the degree of liaison between its access service provider and the relevant production teams in ITV and elsewhere, in particular thanks to regular software upgrades, further knowledge sharing within the team, and increased monitoring and feedback. As a result, 32% of all pre-recorded programmes subtitled live between January and June 2014 had a full script prepared in advance, which was keyed out in block subtitles, compared to 11% in the previous six months. This included, for the first time, regional political programmes. Block subtitles are also used on national and regional news bulletins where scripts are available in advance and where VT items are repeated throughout the day, comprising 30-50% of the total subtitling provided on news programmes on ITV.
- c) ITV has experimented with the use of block subtitles on unscripted live programmes; however, following initial viewer feedback on these trials, it went back to using scrolling subtitles to minimise latency.

ITV also reported that whenever live scrolling subtitles are used, they are automatically repurposed by the subtitling software into block format for future use (e.g. later news bulletins).

- d) Channel 4 said that in the last six months it has increased the instances of block subtitling on its live programmes, particularly on its short news bulletins at noon, which feature very little live or unscripted content, enabling subtitlers to access newsroom running orders in advance.

Channel 4 informed us that it has also increased the use of block subtitles on repeats of programmes which were subtitled live due to fast turnaround – such as *Gogglebox* and *Embarrassing Bodies Live From The Clinic*, and said that it

expects this provision to continue to increase with the help of new technological solutions.

- e) Channel 5 reported that all of its news programmes since May have used block subtitles for scripts and news packages prepared in advance. Scrolling subtitles are used only for sections that are not pre-scripted (for instance, interviews or live updates).

Live subtitles on Channel 5 live programmes – such as *Big Brother* – are reversioned and keyed out in block format for repeats.

- f) Sky said that it has moved to a new software system, which allows it to test the use of block subtitles on live programming. Block subtitles are used on Sky News and Sky Sports News whenever scripts are available to the subtitle provider in advance.

Regarding the reversioning of live scrolling subtitles into block subtitles for later use, Sky said that it uses block subtitles on the majority of repeats of sports events that it covers live, comprising approximately 40% of live subtitled output across its main sports channels.

- 6.3 These are welcome developments. However, it will be important to tackle the problems seen on some of the samples of news programmes measured in the recent exercise which included extracts combining live and pre-prepared subtitles. The external reviewer noted that the availability of scripts for some portions enabled synchronised verbatim subtitles to be produced. However, there were difficulties at the junction between live and pre-prepared subtitling, which often involves a considerable delay in the live part or an excessive speed when live subtitles give way to pre-prepared subtitles so that they can catch up with speech. One suggestion was that pre-prepared blocks could be introduced more gradually with some delay until they are completely synchronous.

Annex 1

External review of measurements



Pablo Romero-Fresco

Inma Pedregosa

22 September 2014

Live Subtitling: A Qualitative Analysis of the Second 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 round of measurements can be found [here](#).

For this second round of measurements, the external reviewers have analysed **a total of 72 ten-minute clips** from 72 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 from Channel 5, ITV News, Britain's Got Talent and Loose Women from ITV and Sky News at Six and Soccer AM from Sky. In total, the analysis comprises twelve hours of live TV material including approximately 124,000 words and almost 19,000 subtitles.

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, both the first and this second round of measurements 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. Only two out of the 72 programmes analysed, despite having accuracy rates of just over 98%, were found to fall below the required quality standard on the basis of the other elements in the subtitles, namely technical problems. 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.

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 98.3%. This is a very similar average accuracy rate to the one obtained in the first round of measurements, but the picture has somewhat changed. Whereas in the first analysis one fifth of the total samples consisted of very good examples (over 99% accuracy) and another fifth was made up of programmes that fell below the required quality threshold, in this second analysis one fourth of the programmes is very good and one fourth is substandard.

In general, the quality of the subtitles has increased, which is very much to the credit of the subtitlers, especially given how challenging some of the programmes included in this sample are, and to the use of pre-recorded subtitles. Yet, on several occasions, the subtitlers' work has been let down by a series of **technical issues** that account for the increase of substandard programmes in this sample. These issues have resulted in subtitles being out of place, garbled, linked to the next ones or interrupted for up to 40 seconds and even over one minute. Needless to say, this has a significant impact on the viewers' comprehension and means that otherwise good subtitles can no longer provide access to some of the programmes in this sample.

The use of more 100% accurate **pre-recorded subtitles** for news and entertainment programmes is welcome and has resulted in an improved overall quality: more programmes with accuracy rates over 99% and lower delays. However, this new hybrid mode has also caused significant discrepancies between the average quality of these programmes and certain instances where this quality is seriously affected by sudden drops in accuracy or by peaks of speed and delay. News and entertainment programmes with an average of 99% accuracy rate have been found to combine segments with perfectly synchronised, 100% accurate subtitles with other segments where the live subtitles fall below the required standard and prevent the viewers from being able to follow specific news items.

Particularly important in this sense is the issue of **subtitling speed**. The average speed of the subtitles in this sample is 143wpm. Research in this area shows that this is an optimum speed that allows viewers enough time to read the subtitles and watch the images on the screen

(approximately 50% of the time on the subtitles and 50% of the time on the images)¹³. None of the programmes has an average speed over 180wpm, generally regarded as the highest recommendable subtitling speed¹⁴ and which would normally allow 60%-65% of the time on the subtitles and 40%-35% of the time on the images. However, in many news and entertainment programmes the introduction of pre-recorded subtitles between sections with live subtitles has resulted in speeds of 290wpm, 350wpm and even 460wpm, which are simply not readable for most viewers. These excessive speeds have been found in one fourth of the programmes combining live and pre-recorded subtitles. In an attempt to catch up with the audio after a piece with delayed live subtitles, some subtitlers seem to be cueing in the pre-recorded subtitles too quickly. We would recommend that the subtitling software is set to a maximum subtitling speed (ideally no higher than 180wpm-200wpm) so that the subtitles can be read in full and 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.

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 imperative to avoid drops in quality and high speeds, which have a serious impact on the viewers' comprehension, thus defeating the purpose of introducing this hybrid mode.

As in the first round 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 **entertainment programmes** analysed here feature the lowest average speech rate (156.6wpm, exactly the same as in the first round of measurements). They 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, since the spontaneous dialogue often features hesitations, false starts and redundant comments. The combination of pre-recorded and live subtitles enables the subtitlers to keep more information from the audio in the subtitles (the average editing rate is 15.77%, as compared to 20% in the previous analysis) and to obtain the highest median accuracy rates of all three genres (98.91%, 0.11% higher than in the previous measurements). Half of the programmes have very good or excellent subtitles and only two programmes fail to meet the minimum requirements in terms of quality due to technical issues.

The **news programmes** in the sample feature a higher average speech rate (176wpm) than entertainment programmes and they normally have more meaningful content, which is 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 challenging. Given the importance of content in this genre, subtitlers seem to be making a significant and commendable effort to edit less than in entertainment programmes (an average editing rate of 13.3%), managing to keep the accuracy rate above the required threshold (98.77%, 0.17% higher than in the first measurement). One third of the examples analysed may be considered as very good or excellent but a few programmes (13% of the total sample) do not meet the minimum requirements in terms of quality, mostly due to

¹³ See Romero-Fresco, Pablo (2011) *Subtitling through Speech Recognition: Respeaking*, Manchester: St. Jerome.

¹⁴ See Ofcom's "Subtitling –An Issue of Speed" (2005) and BBC's Subtitling Guidelines (2009).

technical issues. The combination of live and pre-recorded subtitles accounts for the improved quality of the subtitles provided for news programmes but it has also caused unacceptable peaks in subtitling speed (ranging from 290wpm to 460wpm). As a result, some news items have not been accessible at all for the viewers. Other problems that have been identified in the news programmes analysed are the lack of speaker identification (either no identification for any of the speakers or no distinction between newsreaders and correspondents, which can sometimes cause confusion to the viewers), subtitles that obscure the speakers' mouths (thus making it impossible for viewers to lip read) and subtitles referring to on-screen quotations that have been displayed at least four to six seconds before the subtitle is displayed. Given the unavoidable delay in live subtitling, it may be necessary to include these quotations in the subtitles whenever possible.

Chat shows are the most challenging genre to tackle as far as live subtitling is concerned, not only because of the speakers' fast speech rates (an average of 180wpm in this sample, with peaks of 271wpm) but also because of the presence of overlapping speech that is essential to understand the humour in the programme. The subtitlers have thus been forced to edit an average of 35.17% of the original content. Unlike in the first round of measurements, where the chat shows just managed to reach the accuracy threshold (a median of 98.15%), in this second round they have fallen short, with a median accuracy rate of 97.93%. Whereas 16% of the programmes have good subtitles, 54% do not meet the required standard. In some cases this is the result of the above-mentioned technical issues, while in many others it is simply due to the intrinsic challenges presented by the programmes. In some programmes, editing could have been more effective if subtitlers had managed to leave out unimportant comments, focusing instead on those remarks that carry the comic weight of the programme. For the many occasions on which this is impossible, some programmes have used the symbol (...), also common in other countries such as Canada, so that viewers know that some information is missing in the subtitles. Indeed, one of the most important factors regarding viewers' comprehension is whether or not they are aware that an error (or an omission) has occurred, which could be made clear with the use of this symbol or any other carrying the same meaning. In any case, given the difficulty involved in subtitling these chat shows, it seems that the only way to ensure that they are fully accessible is to have the scripts available before the programme, thus allowing a combination of live and pre-recorded subtitles.

The overall **latency** of the programmes analysed is 5.7 seconds, 0.4 seconds higher than in the previous analysis. The delay of the subtitles depends largely on the availability and use of pre-prepared scripts, 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 average delay has been reduced to 3-4 seconds and on some occasions even to 2-3 seconds. The programmes or sections of programmes that only contain live subtitles produced by respeakers seem to have an average latency of 6 seconds, with very good segments of 5 and even 4 seconds but also peaks of 7 to 9 seconds. In programmes or sections with many speakers, fast speech rates and overlapping speech, the average delay for these live subtitles is 7-8 seconds, with good segments of 6 seconds and peaks of up to 10 seconds. 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.

In the sample analysed, 68% 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 32% 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, 71% of the errors are caused by incorrect editions and 29% by misrecognitions. Entertainment programmes, where speech rates are slower, contain 68% edition errors and 32% recognition errors. Finally, news programmes feature 62% edition errors and 38% recognition errors. This relative increase in recognition errors 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 – would help improve accessibility for the viewers.

As far as the **seriousness of the errors** is concerned, when it comes to the **news**, 80% of the errors are minor (i.e. they do not prevent the viewers from following the content of the programme), 18% are standard (they trigger confusion or cause full factual omissions) and 2% are serious (they introduce misleading information). The evolution of these data with regard to the first round of measurements can account for the improvement in the quality of some of the news programmes included in the present sample. Whereas the type of recognition errors has remained similar (76% minor, 20.5% standard and 3.5% serious in the first analysis vs. 75% minor, 21% standard and 4% serious in the second analysis), edition errors have changed from 74% minor, 25% standard and 1% serious to 85% minor, 14.9% standard and 0.1% serious. In other words, the subtitlers have made a bigger effort to keep up with the news presenters and, whenever something needed to be omitted due to, for instance, time constraints, they have managed to omit mostly minor units as opposed to full sentences. This has increased the overall accuracy rate and improved the quality of the access to news programmes for viewers with hearing loss. In **chat shows**, the fast speech rates and the overlapping interventions of the speakers force the subtitlers to rush and to omit more information. As a result, 63% of the errors are minor and as much as 34% are standard. In other words, one in three errors causes the viewers to lose the thread of the programme.

As far as live **corrections** are concerned, in some cases the correction of minor errors led to segments where the overall quality of the subtitles was affected. Unless there is plenty of time, subtitlers are advised to disregard minor errors (as correcting them only adds delay to subsequent subtitles in cases in which comprehension has not been affected in the first place) and to correct serious errors (those that introduce misleading information) or even standard errors if there is the possibility and time to do so.

The use of colours for **speaker identification** was normally consistent and effective. However, there were also programmes with no speaker identification (which has a significant impact on the viewers' comprehension) and others where the colour green proved problematic, as it is less legible and more tiring to read than white, yellow or cyan. Some news programmes made no distinction between newsreaders and correspondents, all of whom were subtitled in white. In other cases they were subtitled in yellow, but it was not clear whether this was by accident or by design. Given that subtitlers cannot always manage to include the name of the correspondent in the subtitles, it can be very difficult to know who is speaking, which is very confusing for the viewers. In some cases, the use of a new line for the second speaker, as applied in a few of the programmes included in this sample, can be useful in this regard. Finally, some subtitlers used commas for direct speech (i.e. He replied, I have

nothing else to say), which is a good solution given how difficult it is to introduce quotation marks while trying to keep up with the speakers.

3. Conclusions and recommendations

The quality of the live subtitles analysed in this report was good, with an average accuracy rate of 98.3%. The subtitles, as produced by the live subtitlers, were better than in the previous round of measurements, but due to a series of technical issues during transmission, this improvement has not necessarily materialised in a better overall accuracy rate or in a better experience for the viewers.

The use of pre-recorded subtitles alongside live subtitles is welcome and has led to an overall increase in accuracy and a reduction of delay. However, this new hybrid mode has also brought about specific drops in accuracy and peaks of subtitling speed that are problematic for the viewers. Subtitlers now have to switch between live respeaking (in the case of respoken subtitles) and pre-recorded cueing, which is very challenging and does not allow the speech recognition software to have the continuity that is necessary for its optimum performance. Further training on this is likely to improve accuracy results.

Perhaps more problematic are the peaks in subtitling speed found in one fourth of the programmes combining live and pre-recorded subtitles. In an attempt to catch up with the audio after a piece with delayed live subtitles, some subtitles are being cued too quickly, with speeds of 290wpm, 350wpm and even 460wpm, which are not readable for most viewers. It may be useful to set the subtitling software to a maximum subtitling speed (ideally no more than 180wpm-200wpm) so that the subtitles can be read in full and 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.

It is important to ensure that these drops in quality and peaks of speed are avoided so that the current transition from fully live output to the combination of live and pre-recorded output can materialise in improved quality and better access for the viewers. The results are certainly encouraging, as shown by the increased overall accuracy rates obtained in both news and entertainment programmes in this sample. As for chat shows, the low scores obtained here may be explained both by the above-mentioned technical issues and by the very nature of the programmes, which are extremely difficult to subtitle live with no prior information available for the subtitlers. The same goes for pre-recorded entertainment programmes involving constant switching between scenarios and speakers. If they are broadcast with live subtitles, even if they are excellent, the unavoidable 5- or 6-second delay makes it almost impossible for the viewers to identify who is speaking.

To conclude, this second round of measurements has shown further evidence of very good work by the subtitlers to provide access for challenging programmes and of the need to control the subtitling speed in the new hybrid mode, supervise technical glitches in transmission and make available, whenever possible, the scripts for chat shows, entertainment and news programmes.

Annex 2

Additional Charts

- A2.1 The charts in the main body of the paper compare the median results of the measurements of accuracy, latency, speed, and edition rate between the first and second sampling exercises. The median is the midpoint of the measurements observed in the sample, such that half of the measurements fall above it and half of them fall below it.
- A2.2 The charts in the first report included the minimum and maximum values for all the measurements, as we believe that these statistics provide valuable information for the assessment of the quality of live subtitling. This Annex includes charts showing a comparison of the minimum, median and maximum results for each dimension of quality between the first and the second rounds of measurements. The charts in this Annex follow the order of the charts in the main body of the paper.

Accuracy

Figure 13: accuracy rates in samples drawn from news bulletins
(higher bars represent a better viewing experience)

NER Score

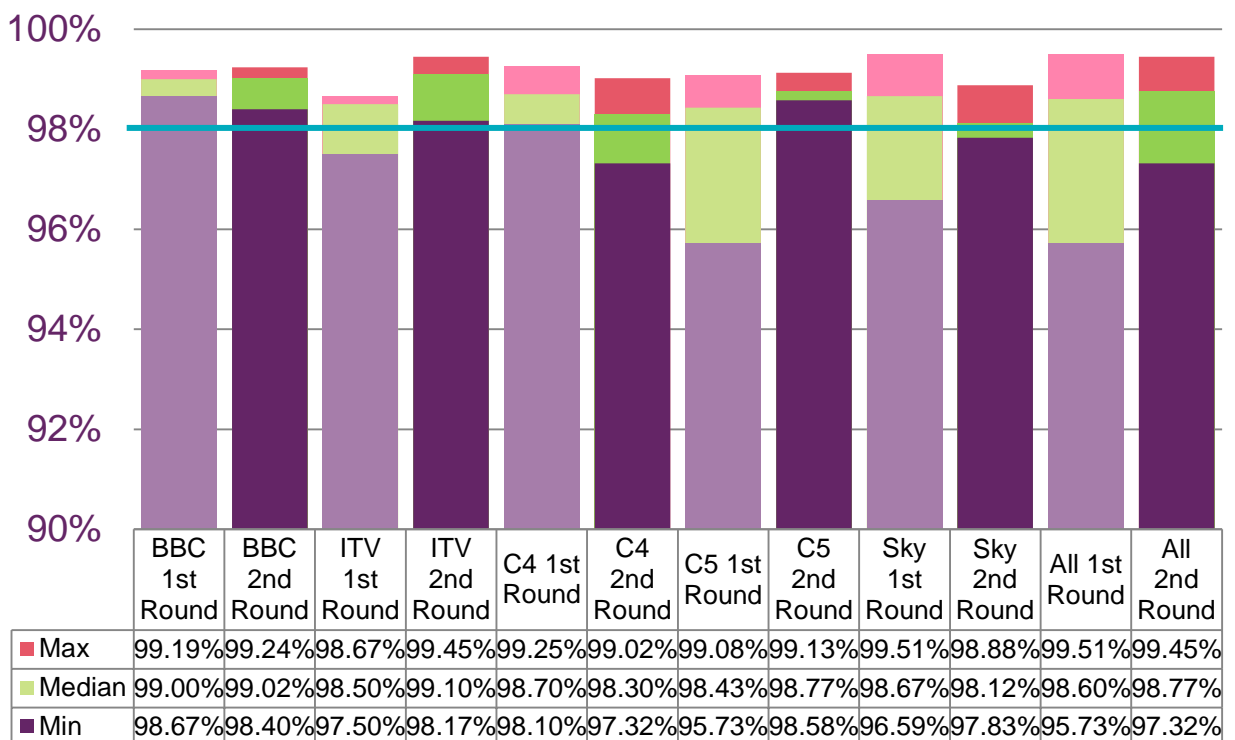


Figure 14: accuracy rates in samples drawn from chat shows
(higher bars represent a better viewing experience)

NER Score

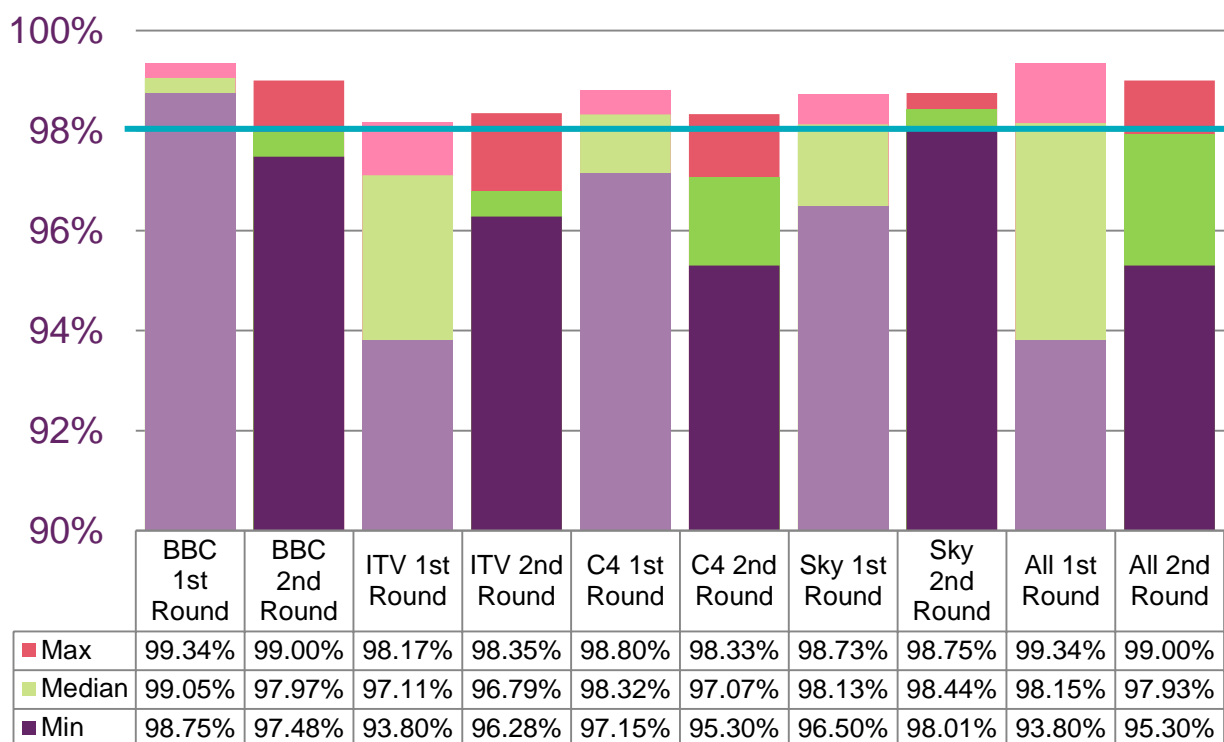
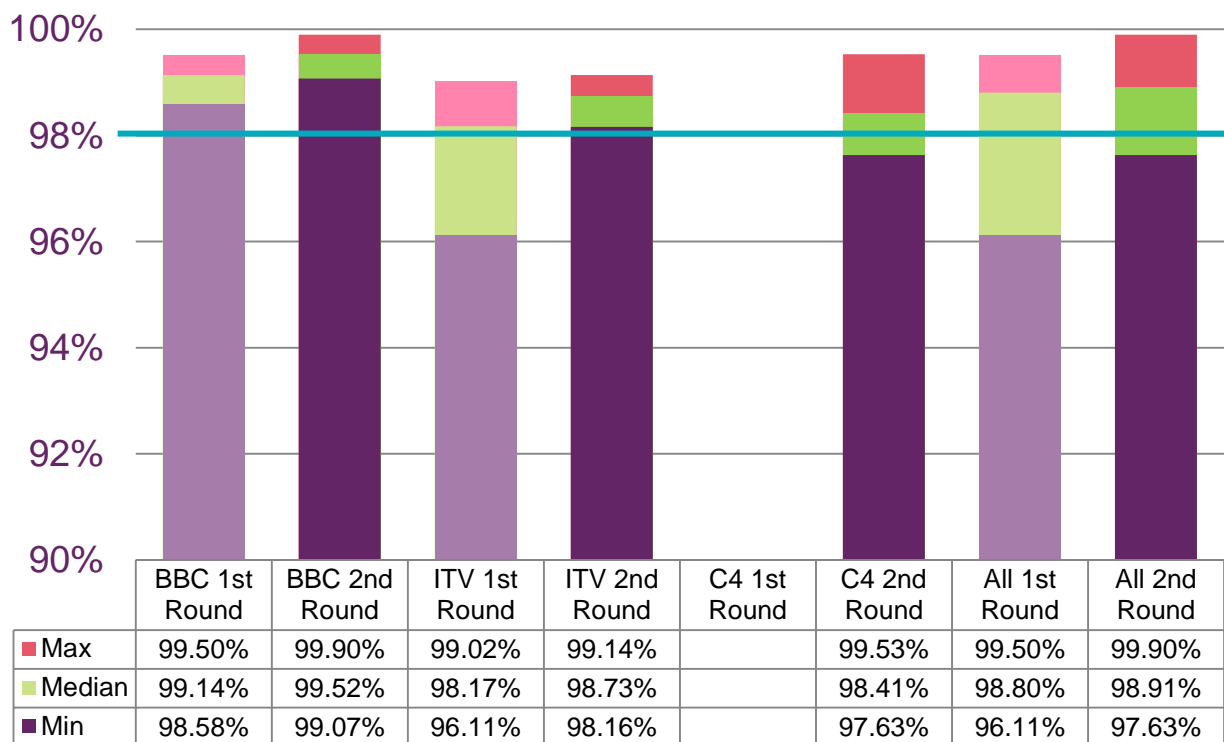


Figure 15: accuracy rates in samples drawn from entertainment shows
(higher bars represent a better viewing experience)

NER Score



Latency

Figure 16: latency in samples drawn from news bulletins

(higher bars represent an inferior viewing experience)

Seconds (")

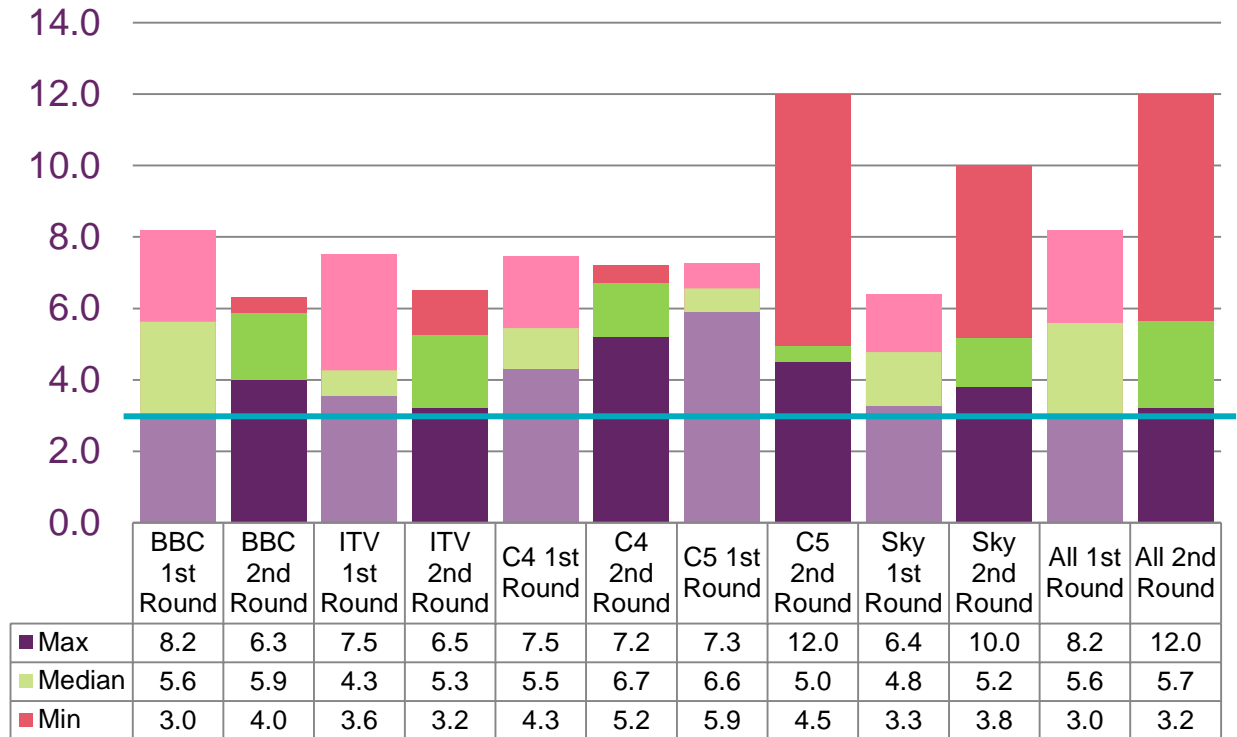


Figure 17: latency in samples drawn from chat shows
(higher bars represent an inferior viewing experience)

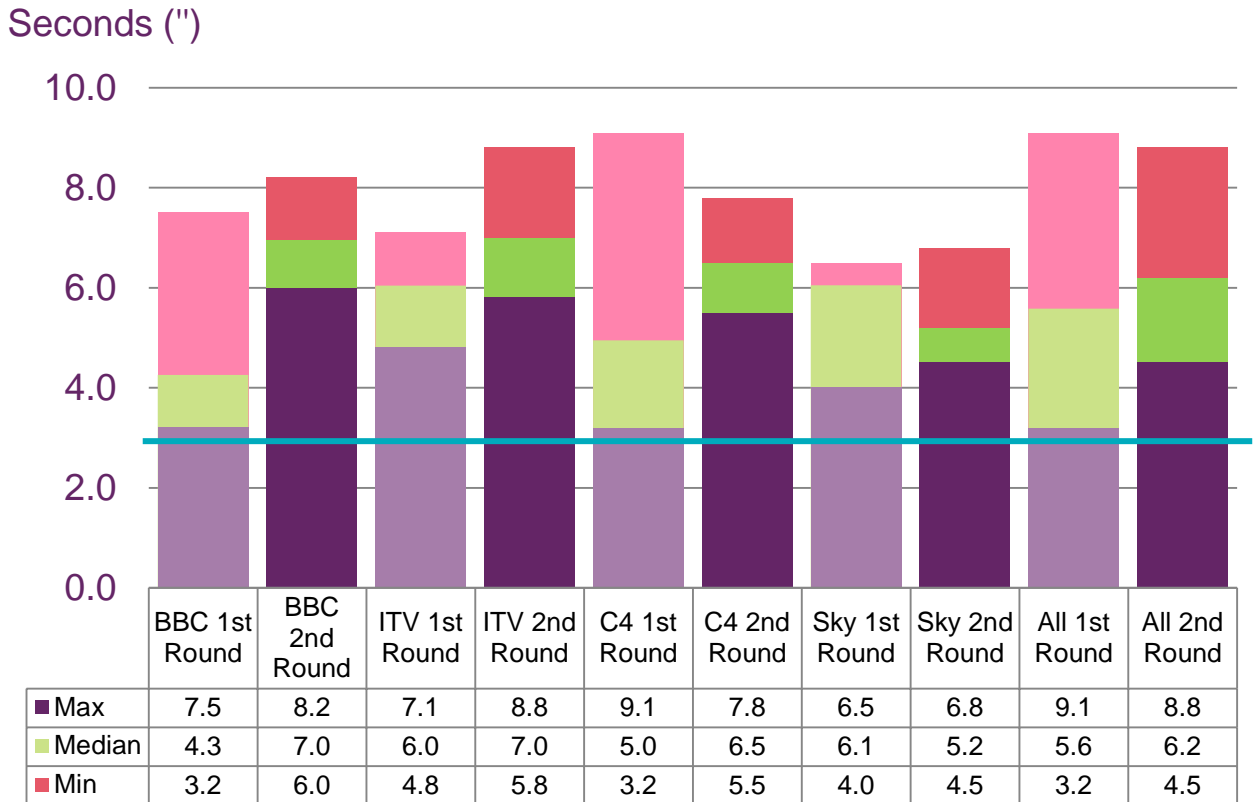
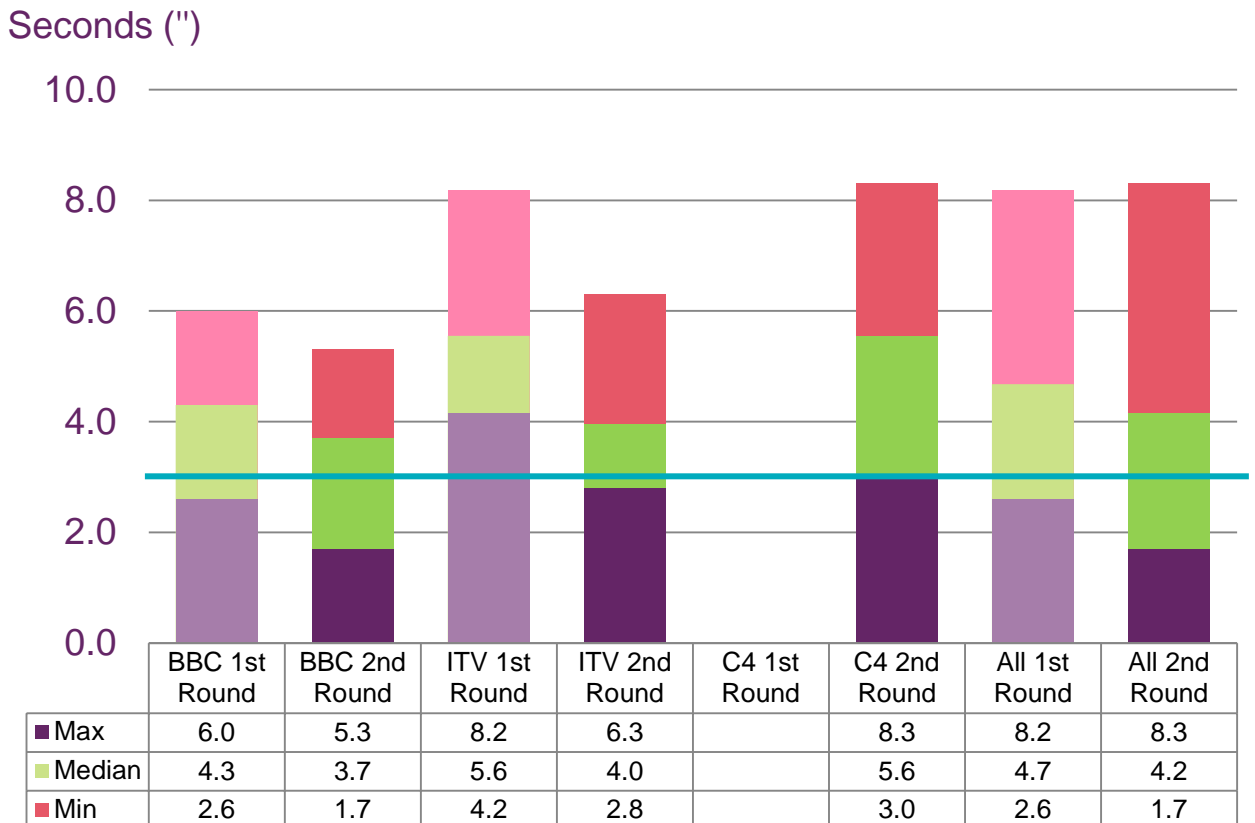


Figure 18: latency in samples drawn from entertainment shows
(higher bars represent an inferior viewing experience)



Speed

Figure 19: speed of subtitling in samples drawn from news bulletins

Words per Minute (wpm)

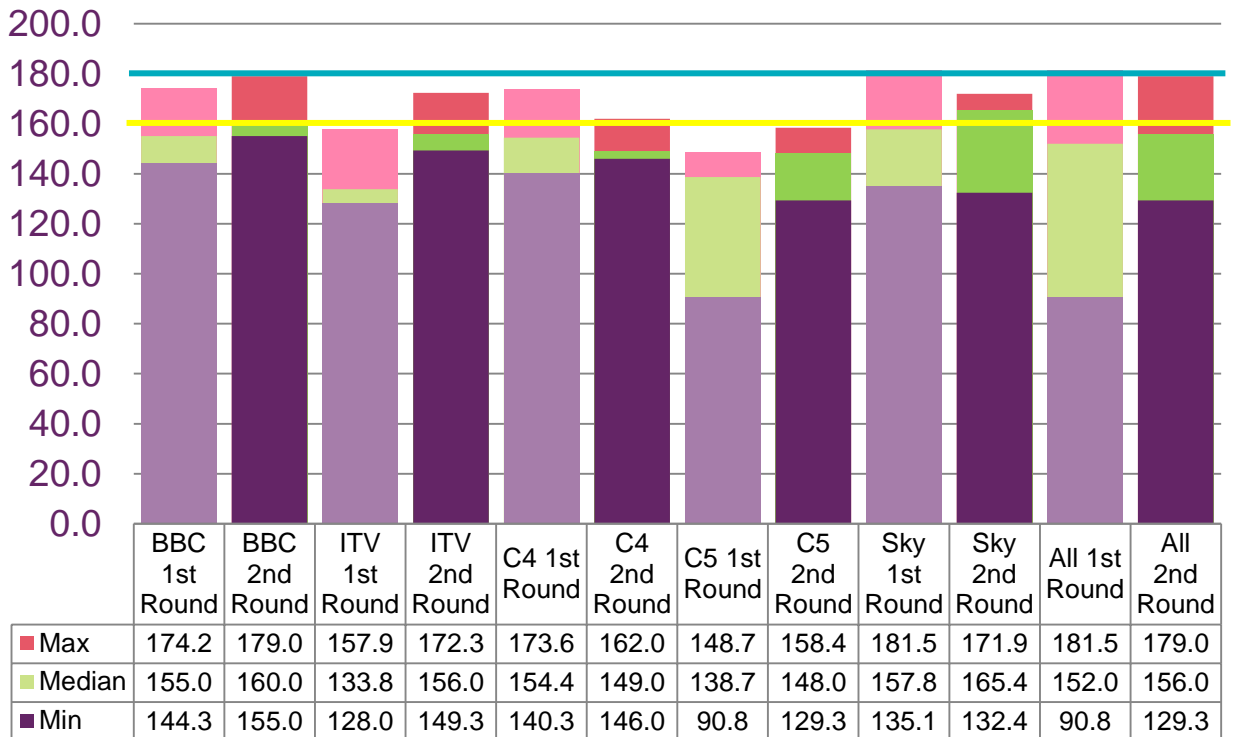


Figure 20: speed of subtitling in samples drawn from chat shows

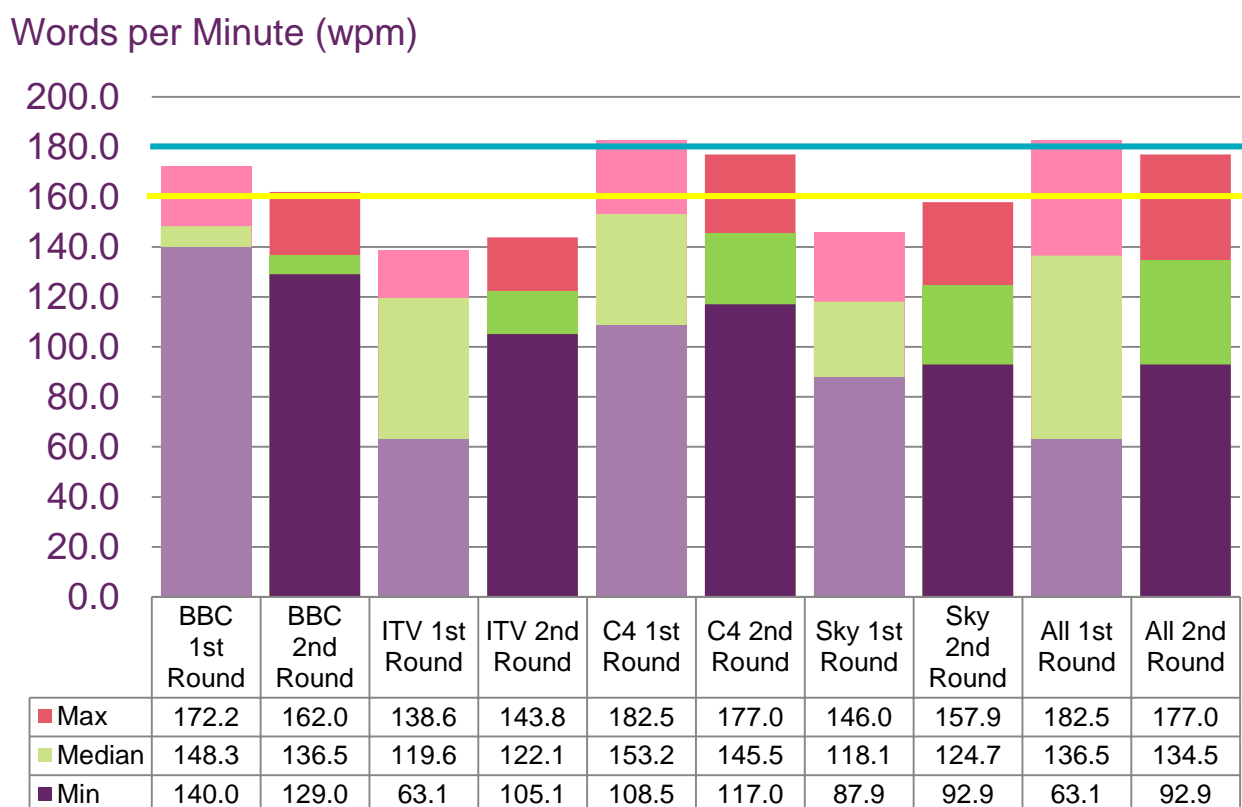
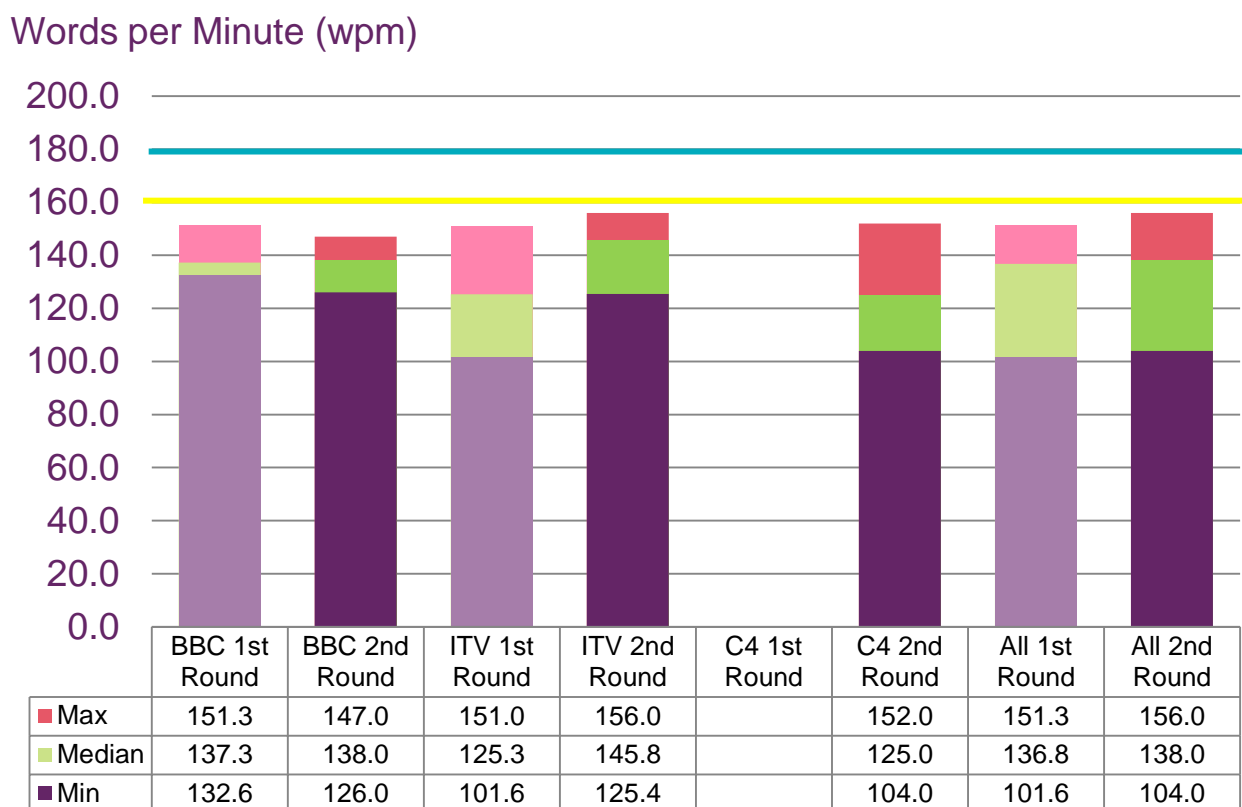


Figure 21: speed of subtitling in samples drawn from entertainment shows



*the first round of measurements did not include samples of entertainment programming shown by Channel 4

Edition rate

Figure 22: edition rates in samples drawn from news bulletins

Edition Rate

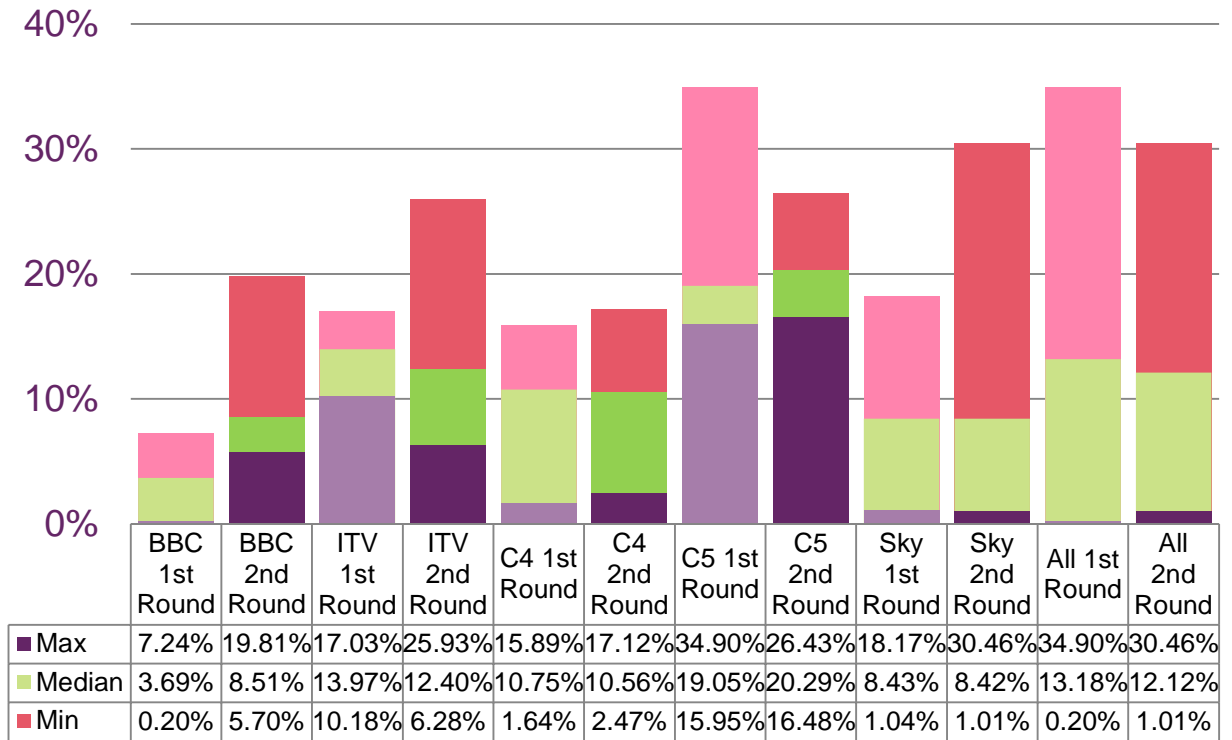


Figure 23: edition rates in samples drawn from chat shows

Edition Rate

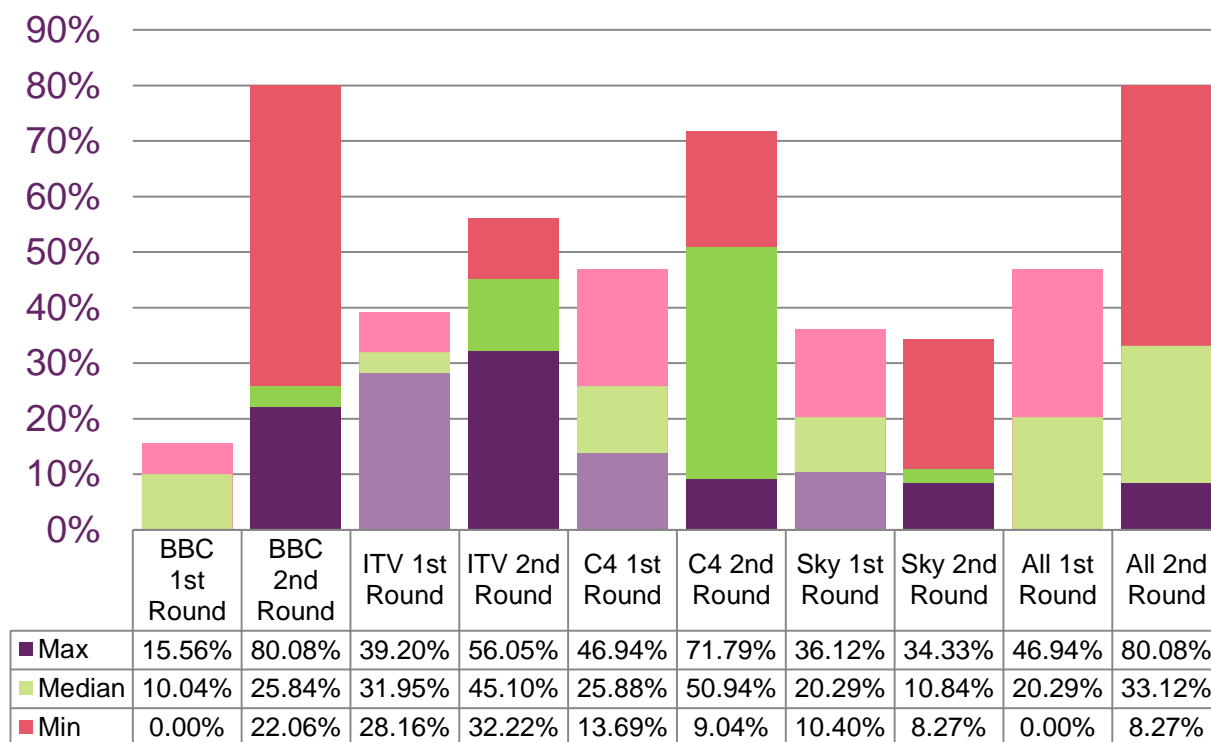
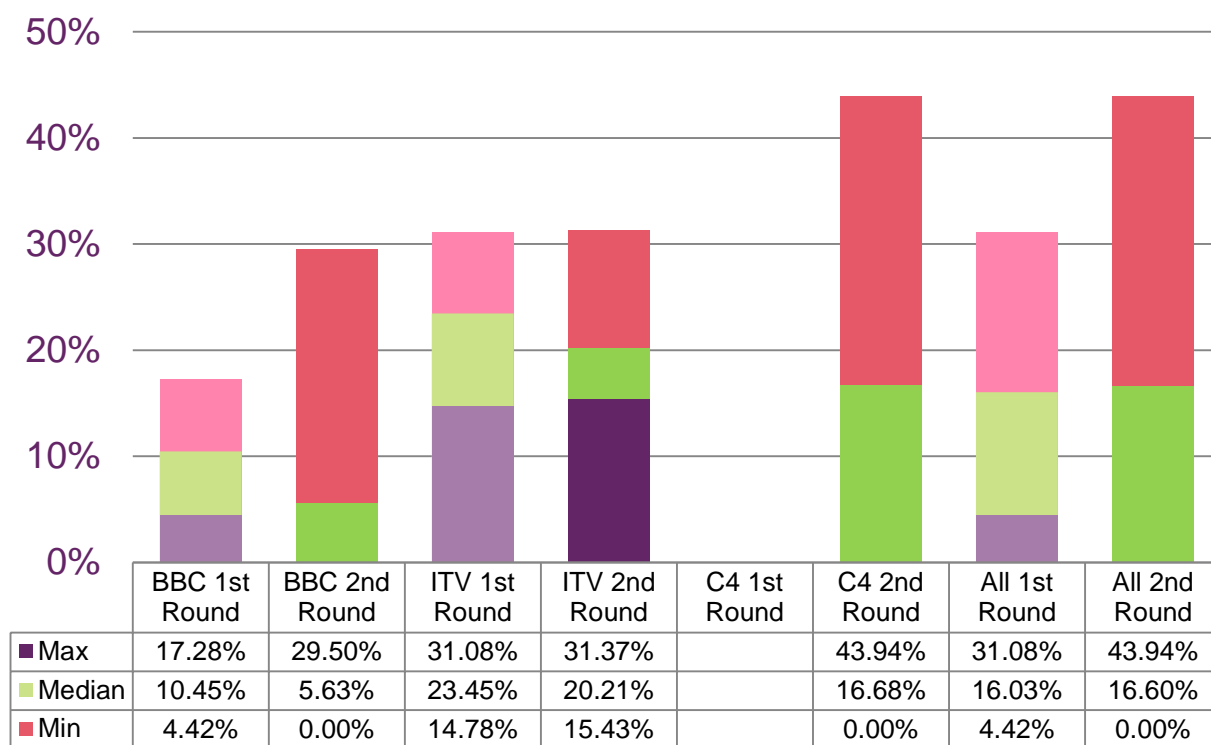


Figure 24: edition rates in samples drawn from entertainment shows

Edition Rate



*the first round of measurements did not include samples of entertainment programming shown by Channel 4