IBC2024

# PROJECT TIMBRE: HOW WELL DO MOBILE NETWORKS WORK FOR LIVE AUDIO STREAMING?

A. J. Murphy, S. D. Elliott

British Broadcasting Corporation, United Kingdom

## ABSTRACT

In Project Timbre, data is being collected from around twenty standard, off-the-shelf mobile phones to examine the performance of live audio streaming over today's mobile networks. The project focusses on assessing the real-world quality of experience and investigates how this correlates with existing definitions of mobile coverage, which may be more general in nature.

The analysis presented uses a statistical approach based on the collected data to examine the probability of a given level of service, much as is done today for broadcast networks. It shows that there can be a marked difference between so-called 'signal coverage' and 'service coverage' and has identified the 'Time to First Byte' as a key metric of interest for live audio distribution over mobile networks.

This paper also makes the case for using better data to enable continued dialogue between content providers, the mobile industry and regulators to both optimise live audio streaming and to better inform listener expectations.

## INTRODUCTION

Audio apps from broadcasters give access to the widest possible range of both live radio stations and on-demand content. Listening live and on the move, for example as a pedestrian or in a car, is almost certainly delivered by mobile networks.

But how well do mobile networks work for live streaming?

In the European Union, 4G (LTE) coverage is already termed 'near-universal' [1], reaching 99.8% of all households. However, this headline figure can belie the real-world experience when further away from home.

In the UK, 93% of the landmass and 98% of motorways & main ('A') roads have 4G coverage from at least one Mobile Network Operator (MNO) [2], reducing to 71% and 69% respectively from all operators. Schemes such as the *Shared Rural Network* [3] are expected to see this rise to 95% of the landmass from at least one operator (84% from all operators) by the end of 2025.

But what does 'coverage' mean in the real world for listeners' Quality of Experience (QoE) in the context of audio apps?

Illustrating a wider held desire to find out more about listeners' QoE over mobile networks, the UK Government's 2022 Digital Radio and Audio Review [4] made recommendations for further work on mobile networks and radio streaming, specifically:

> *R32: Industry should work closely with Mobile Network Operators to promote the build-out of robust mobile data networks (5G) and deliver on-demand, streamed listener services focused on in-car listening.*
>
> *R33: … radio broadcasters, transmission providers and Ofcom should initiate a programme of field-testing and trials to review and validate the … findings on 4G/5G coverage. The results of this testing should be discussed with Ofcom to ensure they include in their Connected Nations reporting, a measure appropriate for reliable radio/audio streaming.*

*Project Timbre* is starting to investigate these issues by focussing on the real-world Quality of Experience (QoE) for listeners using the BBC's own audio product, *BBC Sounds*. It concentrates on live radio as the most challenging use case, since this requires a constant, reliable internet connection to successively download audio segments as they are created i.e. it is not possible to download live audio via another means before it has happened. However, the concepts being explored can also apply to streaming of on-demand content.

This paper sets out the work carried out in the project so far and presents some preliminary thoughts and findings as well as ideas for future work.

## QUALITY OF EXPERIENCE

Broadcasters have a high confidence in the QoE delivered by conventional broadcast radio transmitter networks. These are downlink-only and dedicated to live audio delivery, being built for the specific requirements of broadcast. They deliver high audio quality with low latency and few interruptions i.e. 'broadcast quality'.

Mobile networks, in contrast, are multi-purpose and built to satisfy a broad range of simultaneous uses and requirements. These bi-directional IP networks enable the full range of content – both live and on-demand – to be made available, while offering the potential for new experiences and personalisation. Here, it is harder to have confidence in the QoE since it varies depending on many factors such as time of day (e.g. rush hour vs. night-time) and location (e.g. city centre train stations vs. countryside). It is also affected by the complex interaction between the dynamic nature of the network and the response of the algorithms in the playback client to those varying characteristics.

However, it can be anticipated that listeners will expect a similar quality of experience for streamed live audio as that provided by broadcast.

A plethora of sources of mobile coverage information is already available to the public, such as coverage checkers, maps based on predictions and signal strength thresholds or speed-test data. However, the variability outlined above means that these existing sources of information may describe the *signal coverage* but do not always accurately convey the QoE for audio streaming or indeed the *service coverage* for any specific application. This can make it difficult for the user to know what to expect.

In comparison to broadcast networks, the bi-directional nature of mobile networks offers an opportunity to collect feedback on performance and to harness data to better understand the real-world performance of services in relation to predictions of mobile coverage as well as to optimise products like BBC Sounds to improve QoE in the mobile environment.

## DATA COLLECTION

In *Project Timbre*, an augmented, private prototype of the Android version of BBC Sounds has been developed. This has been deployed internally to engineers who are using it on standard, off-the-shelf handsets to collect real-world QoE data over mobile networks as they are today. With no need for special test equipment or targeted drive-test campaigns, a significant amount of data has been cost-effectively logged from the everyday trips of a small number of engineering staff (Table 1). While the phones are not calibrated, they capture real-world experience data, which is used to explore variations due to location and time of day and to identify metrics having the greatest impact on QoE. The data also helps to enable better collaboration with wider industry to improve QoE.

| MNO | Hours | Roads, km (% of total) | | Rail, km (% of total) |
|---|---|---|---|---|
| | | Motorways | A + B | |
| A | 3,300 | 2,900 (77%) | 13,800 (26%) | 1,850 (14%) |
| B | 1,950 | 2,200 (59%) | 7,450 (14%) | 1,250 (10%) |
| C | 3,350 | 2,500 (66%) | 10,500 (20%) | 1,350 (10%) |
| D | 1,500 | 1,500 (40%) | 5,000 (10%) | 1,000 (7.5%) |

Table 1 – The long and winding roads: length of transport routes surveyed (and as a percentage of UK total) over 32 months


Figure 1 – Data collection in Project Timbre

Data is collected at various points in the distribution chain (Figure 1).The live radio stream originates at the BBC, where it is encoded and packaged as a sequence of MPEG-DASH segments. These are delivered using HTTP via, Content Delivery Networks (CDNs) and a mobile network to the BBC Sounds app running on a smartphone (or, in future, perhaps a connected car). The sequence of discrete audio segments is reassembled into a continuous stream of audio to be played back to the listener. Four categories of data are being collected:

1. **QoE metrics**: These are derived from the audio playback client (the BBC's in-house Standard Media Player, which in turn is built on ExoPlayer);
2. **Audio Delivery metrics**: These are provided by the underlying HTTP library used to download each of the audio segments that constitutes the live stream;
3. **Network Quality metrics**: The Android Telephony APIs enable the collection of detailed information about the mobile network in use, its signal strength, signal quality and the primary serving cell identifier; and

4. **CDN View**: Logs from the CDNs give a server-side view of what is happening with the connection and the delivery of each individual audio segment.

This data is recorded second-by-second and logged against location and time.

The client-side data (1, 2 and 3 above) is collected using Message Queue Telemetry Transport (MQTT) messages that are delivered to a back-end Influx database. Local storage in the client is used to ensure that data is captured everywhere, including in areas with no, or insufficient mobile signal, with messages sent to the database once connectivity returns.

Grafana dashboards monitor and debug the real-time system. A snapshot of this (Figure 2) depicts playback state (top row, map left & graph centre), behind live latency (bottom row, graph centre), a timer to measure the delivery of individual audio segments (top row, map right) and mobile signal metrics (bottom row, map left & graph right).

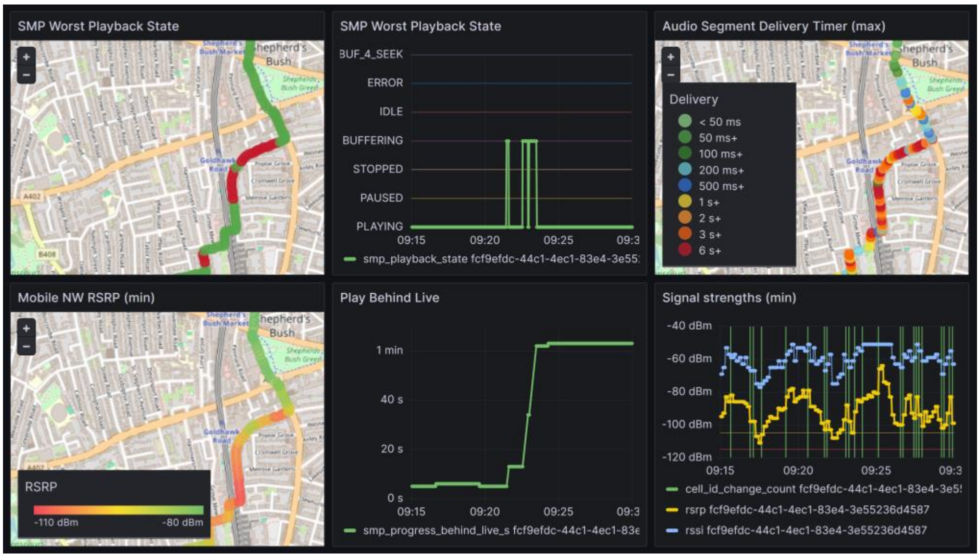


Figure 2 – Real-time monitoring of a single measurement session

## DATA INSIGHTS

The ultimate listener-experienced QoE is defined by a number of factors such as latency, audio bit rate and both the frequency and duration of any audio interruptions. In order to better understand one of these areas, a simple proxy for QoE has been defined as the fraction of playing time ($T_p$) to overall listening time ($T_l$), or Playing to Listening Ratio (PLR):

$$PLR_{[\%]} = \frac{100\ T_p}{T_l} = \frac{100\ T_p}{T_b + T_p} \qquad \{1\}$$

where $T_b$ is the buffering duration when playback was interrupted.

As broadcasters and audiences typically seek uninterrupted playback for extended periods, a suitably high PLR target needs to be considered. The relationship between $PLR_{[\%]}$ and $T_b$ for an hour's listening is shown in Table 2.

| $PLR_{[\%]}$ | $T_b$ (s) |
|---|---|
| 97.5 | 90 |
| 99.0 | 36 |
| 99.5 | 18 |
| 99.8 | 7.2 |
| 99.9 | 3.6 |

Table 2 – $PLR_{[\%]}$ vs $T_b$ for one hour of listening

Figure 3 shows that the $PLR_{[\%]}$ in the beginning of 2024 has remained above 97.5% across all MNOs. However, there is distinct variation over both time and between different networks.

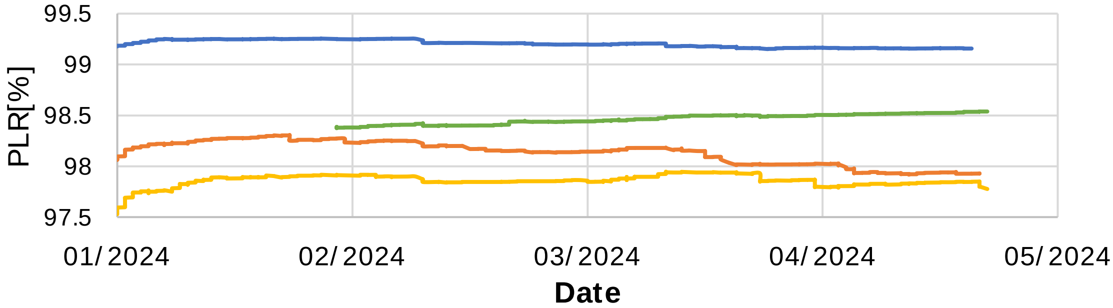


Figure 3 – Overall $PLR_{[\%]}$ for the four physical UK mobile networks, 2024 to date

This high-level view of QoE based purely on a time-series analysis inevitably disguises the variations according to the changing locations and environments of the population of devices, and the network performance experienced while logging. For example, the QoE is often poorer in trains than elsewhere due to greater signal attenuation of train carriages and railway cuttings compared with typical in-car reception.

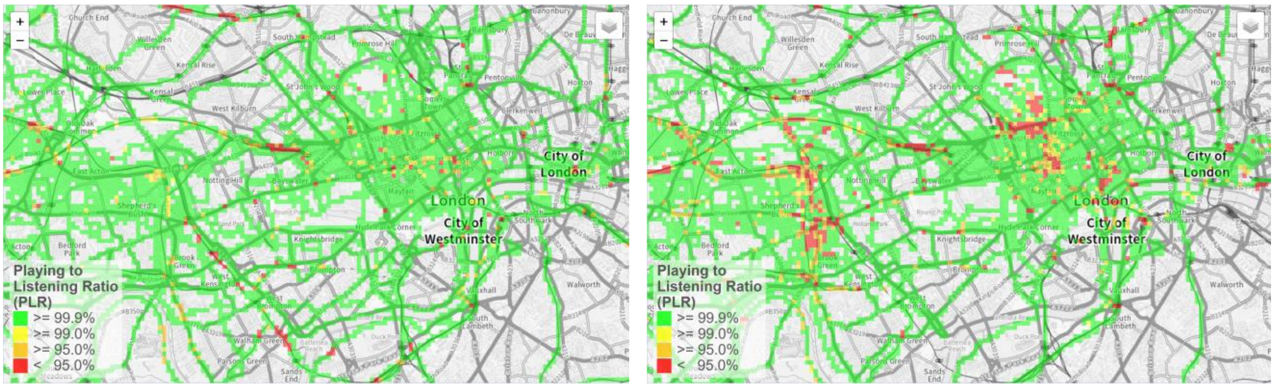


Figure 4 – $PLR_{[\%]}$ variation over location for two different MNOs

To consider location, multiple data points are aggregated into 100m-by-100m pixels. Figure 4 shows the $PLR_{[\%]}$ for two different networks in west London. In general, the QoE has been very high. However, in some locations the QoE has clearly been consistently poorer than in others. Areas of poor QoE common across networks often indicate challenging local topography such as railway lines in cuttings while poor QoE occurring in different places from one network to another are more likely due to differences in network deployments. QoE can therefore vary from location to location and from network to network.

Variations in QoE between different mobile networks are important to be aware of as they create greater complications and potential confusion for both listeners and content providers alike. These are in sharp contrast to the QoE offered by a single broadcast network providing a near-universal service.

When analysing delivery over mobile networks, both temporal and spatial data are therefore needed and can be used to promote realistic consumer expectations and for broadcasters in planning, optimising and promoting their app-based products and services, as well as informing broadcasters around the provisioning and deployment of future experiences.

## QUALITY OF EXPERIENCE AND BUFFERING

While $PLR_{[\%]}$ is a useful proxy for the Quality of Experience, it is hard to measure objectively.

The public version of BBC Sounds normally makes use of adaptive bit rate (ABR) streaming, choosing the most appropriate stream quality and codec depending on its view of a user's network bandwidth at any given time. However, to reduce the number of variables and to concentrate on the delivery of a minimum service, in Project Timbre the bit rate and codec are fixed at the lowest bit rate representation, namely HE-AACv1 at 48 kbit/s.

The audio segment duration on BBC Sounds is 6.4 seconds, chosen as a compromise between minimum latency, compatibility with early playback clients and the need to encapsulate an integer number of audio frames for seamless ABR switching. Since there is no variation of the playback speed, to play live audio uninterrupted – and to keep up with the live edge of the broadcast – the client player needs to request and download successive audio segments every 6.4 seconds, on average.

Fluctuations in the time it takes for each segment to be delivered over the network (the delivery time, $T_{delivery}$) are smoothed out by a buffer in the client app, the length of which effectively sets the averaging period over which the above 6.4 second criterion must be met.

Under 'normal' conditions, where the entire distribution chain (including, for example, CDN, mobile network and client) performs sufficiently well, successive audio segments will typically arrive within the order of tens of milliseconds. Such conditions enable uninterrupted playback with a very short buffer and low resulting latency. However, playback interruptions ('BUFFERING' events) do happen in practice, with the result that, when playback resumes, the listener is subject to increased latency; they don't miss that vital goal being scored, but rather, they hear it later.

Barring any erroneous operation of the BBC Sounds app itself – or being in an area without any mobile coverage – playback interruptions will be the result of excessive delay (i.e. more than 6.4 seconds) in the requested audio segment data being delivered.

The client's audio buffer decreases the chance of a delayed segment interrupting playback, and there is a relationship between the length of this buffer and the tolerance to excessively long delivery times. A longer buffer means that the client can wait longer for a segment to be delivered before an audio interruption occurs. However, this increased resilience comes at the cost of increased latency for the listener; increasing the risk of hearing about that goal from their friends before they've heard it for themselves.

The initial length of the client buffer at which playback starts is a design decision that needs to balance the less desirable aspect of increased latency on the listener experience against the initial resilience to excessive segment delivery times.

Further increases in latency due to playback interruptions allow the client to maintain an even longer buffer. Client algorithms are typically 'greedy', with any increase in latency an opportunity for players to gorge themselves on audio segments up to the latest available at the live edge, subject ultimately only to memory constraints on the device.

As a result, any prior interruptions to playback that result in increased latency for the listener – such as those caused by excessive segment delivery times brought about by poor network coverage – result in a longer buffer and hence more resilience to subsequent segment delivery delays. There is further dependence of buffer length on how the user has controlled playback (e.g. pausing or rewinding).

The net effect is that different users will report different values of $PLR_{[\%]}$ depending on their past behaviour. A single user may even experience a different $PLR_{[\%]}$ in the same location on the return vs. the outward leg of their journey.

In summary, the listener's true QoE is a complicated product of many factors, initial client buffer length, including user interactions, exposure to previous network outages and segment delivery delays that introduce latency as well as the behaviour of the playback client in reaction to those. QoE may even vary depending on the programme content, with for example, someone listening to live sport being most sensitive to latency as outlined above.

Of interest, therefore, are occasions when the delivery time ($T_{delivery}$) of the current audio segment is excessively long since this has the *potential* to cause an audio interruption and hence a reduction in $PLR_{[\%]}$ (Figure 5). It also acts as a common currency, being neither dependent upon, nor impacted by, variation in the instantaneous length of the client audio buffer.

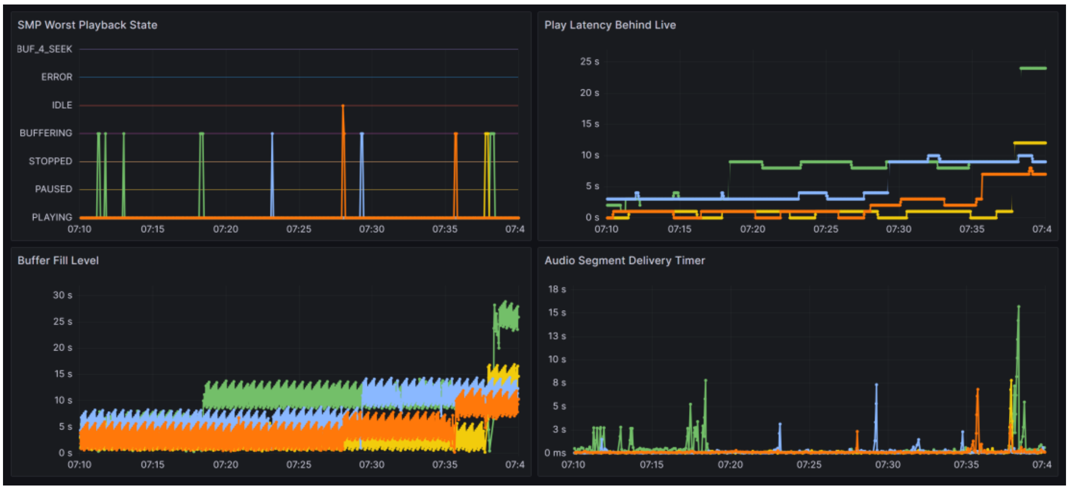


Figure 5 – Impact of $T_{delivery}$ on QoE for four handsets, across four MNOs, same journey

The '*BUF_4_SEEK*' playback state is used to distinguish between buffering that occurs at the behest of the listener (i.e. when changing station, fast-forwarding, etc.) and buffering caused by delay in delivery somewhere in the underlying distribution chain.

Over the course of the thirty minutes depicted here, several 'BUFFERING' instances can be seen to occur, causing pauses in the audio. These are a direct result of certain audio segments having an excessively long delivery time and happen at different times on different mobile networks. The net effect is a spread of buffer lengths and resulting variation in audio latencies from around 7 to 24 seconds across the four handsets.

## SEGMENT DELIVERY TIME IN DETAIL

The time taken for the HTTP transaction for each segment is the sum of two components:

$$T_{delivery} = T_{TFB} + T_{transfer} \qquad \{2\}$$

$T_{TFB}$ is the time taken between the initial request and reception of the first byte of the response from the server (typically a CDN end-point), the so-called Time To First Byte and $T_{transfer}$ is the time taken to transfer the data itself, in this case the audio.

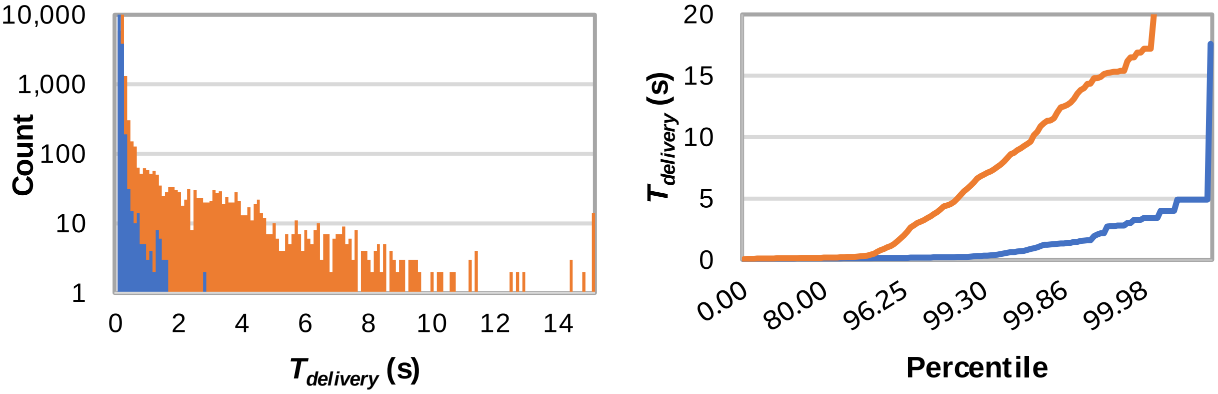


Figure 6 – $T_{delivery}$ (histogram left, cumulative right) in a single pixel, two MNOs, 4G only

Figure 6 takes a closer look at the distribution of this segment delivery time ($T_{delivery}$) of 4G-only data logged for two networks in a single 100m-by-100m pixel. The median (50$^{th}$ percentile) $T_{delivery}$ for both networks is approximately 150 ms; even a very short client buffer would 'mop up' latency of this order, seemingly preventing audio interruptions.

However, $PLR_{[\%]}$ observed for the network depicted in blue exceeded 99% while it was less than 95% for the network depicted in amber. This is due to the amber distribution's long tails extending to the right, revealing that significantly longer latencies do occur. As the 99.8$^{th}$ percentile for the blue network was around 1.2 s, very few audio segments would have been delayed by more than a typical buffer length, enabling a high QoE. On the other hand, long segment delivery times were more common for the amber network, where the 99.8$^{th}$ percentile has been around 11.5 s. These more frequent, longer $T_{delivery}$ values, more often exceeded the buffer length, resulting in more frequent audio interruptions, and lower QoE.

Segment delivery time is therefore a key indicator of QoE, with the 99$^{th}$ or higher percentiles being of primary interest.

Figure 7 shows the distribution of a larger number of segment delivery times across multiple pixels, as the constituent $T_{TFB}$ and $T_{transfer}$ elements. The occasional instances of complete download failure are captured in the bin labelled 'FAIL'.

Somewhat surprisingly, $T_{TFB}$ dominates over $T_{transfer}$ i.e. it often takes longer for the first byte of the data request to arrive than it takes for the rest of the payload to be transferred. This may be due to the small audio segment size but requires further investigation.

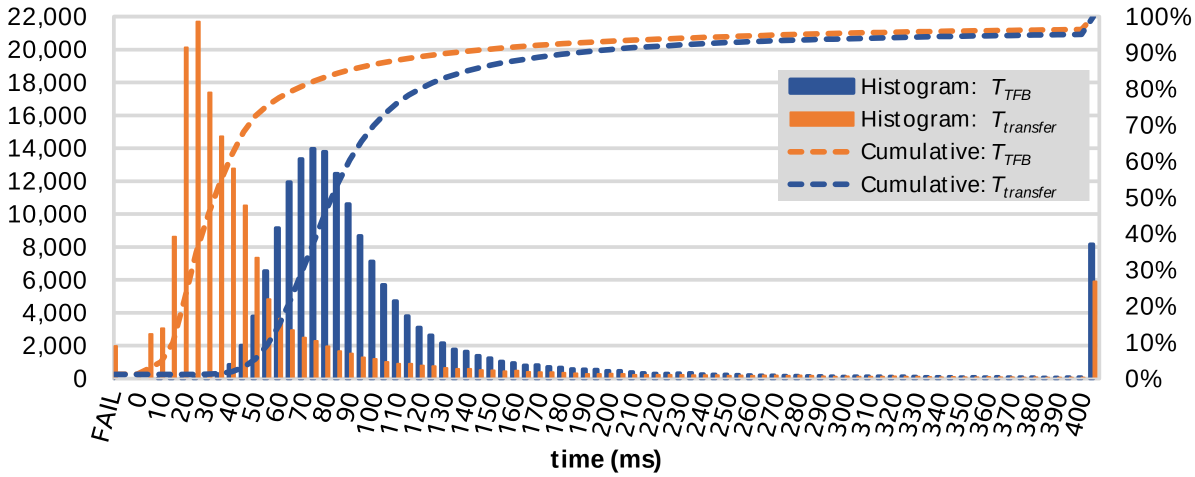


Figure 7 – Distribution of $T_{TFB}$ and $T_{transfer}$ for one MNO

Furthermore, there appears little correlation between them (Figure 8) i.e. longer $T_{TFB}$ does not always imply longer $T_{transfer}$. This may be important for adaptive bit rate services where clients rely on accurate assessments of network bandwidth. Simply measuring transfer time ($T_{transfer}$) may not give a true assessment of the overall segment delivery time for audio and can result in a significant overestimate of connection bandwidth.

Figure 8 – $T_{TFB}$ vs $T_{transfer}$ for one MNO, all data

## NETWORK METRICS AND QOE

Determining the causes of any poor QoE is of interest. In the UK both 2G and 3G remain in use, with the latter being phased out. Both older generations often show long segment delivery times compared with 4G (Figure 9) resulting in audio interruptions.

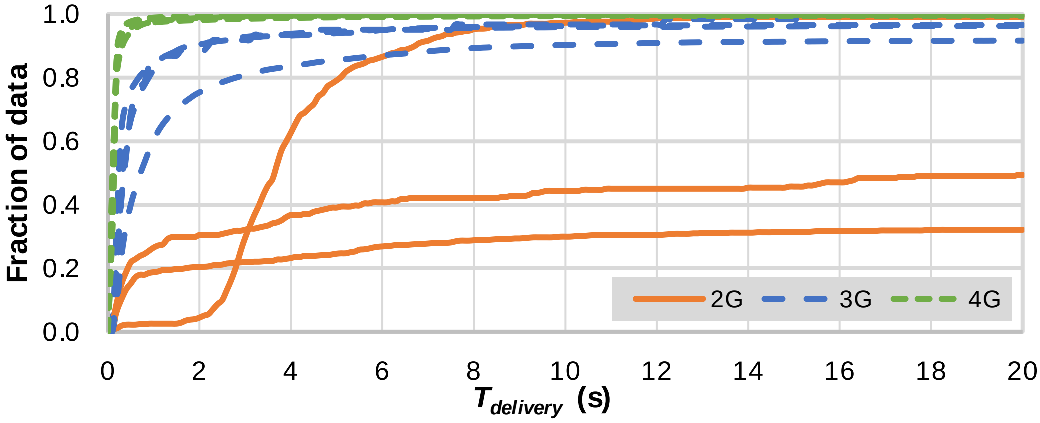


Figure 9 – $T_{delivery}$ for the different generations of mobile technology

Although considerably better than 2G and 3G, unusually long segment delivery times are also observed on the 4G networks. The cause of many of these delays is as yet unclear.

To stream audio, a device must, at a minimum, be connected to a network with sufficient signal. Logging signal strength (RSRP) and network connection status has enabled the identification of locations where either one of these is untrue. Insufficient signal for a network connection has, as expected, been found to be the cause of poor QoE in some locations, particularly in remote rural areas with rugged terrain, but also in urban areas, albeit less frequently. Other locations, however, appear to have sufficient signal yet frequently suffer long segment delivery times. The expectation is that these areas are congestion limited.

Figure 10 depicts the relationship between signal strength and segment delivery times by mapping RSRP against the observed downlink bit rate (calculated over $T_{delivery}$) for two different networks. Although not shown, similar variation has been observed for the RSRQ. It demonstrates that increased signal strength is no guarantee of improved throughput (reduced $T_{delivery}$), especially at lower percentiles. For example, at the 2$^{nd}$ percentile the network depicted in the bottom chart provides a throughput of 132 kbit/s at -60 dBm, compared with a higher throughput of 536 kbit/s at -71 dBm RSRP.

These findings suggest that other factors not captured by these metrics – such as the aforementioned network congestion (in either the up- and/or downlinks), handover delays, and potentially operational issues – contribute to the performance observed.

All that can be stated with any certainty is that signal strength is a necessary but not always sufficient requirement for adequately fast segment delivery times. This underlines the point that correlating the coverage available for live audio delivery with a given QoE – i.e. '*service coverage*' – to maps or data based on signal strength alone – i.e. '*signal coverage*' – appears fraught with error.

Of the other signal quality measures, SINR would be expected to provide the best insight into network performance. However, limitations of the telephony APIs available on Android has not made it possible to log SINR with enough regularity to test the relationship.

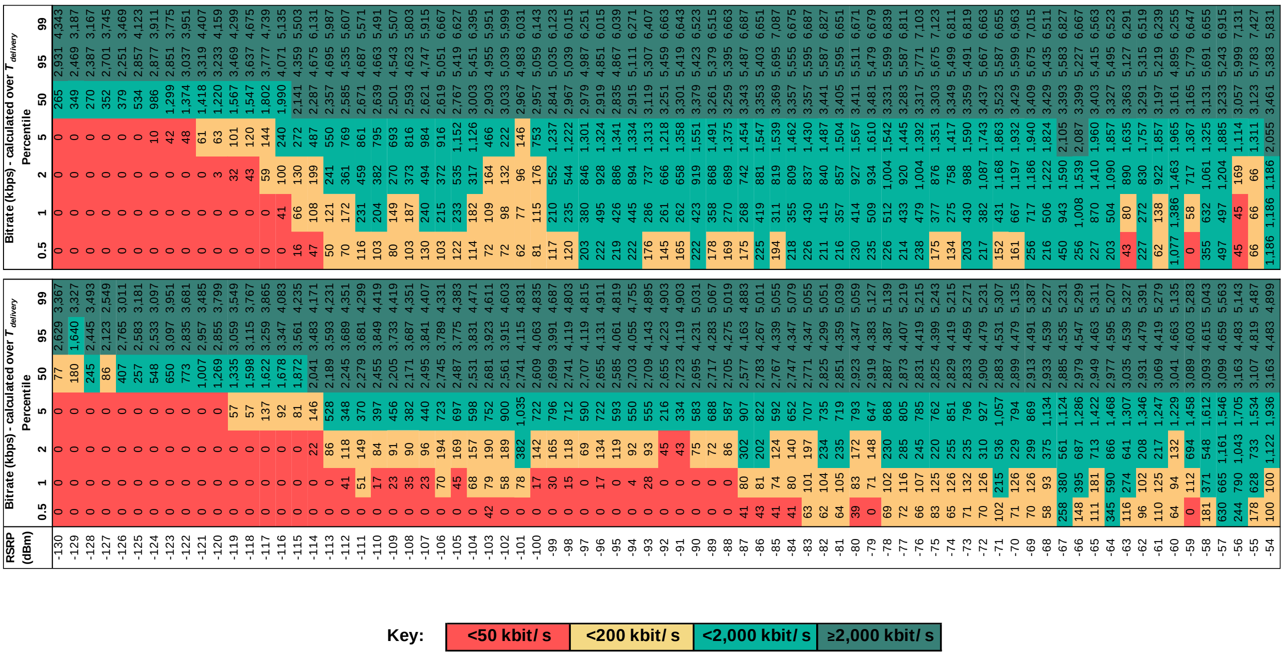


Figure 10 – Measured Downlink speed vs. RSRP for two MNOs, RSRP <= -54 dBm

## CONCLUSIONS

Listeners are consuming more and more content through audio streaming apps such as BBC Sounds. On the move, these are reliant on mobile networks for the delivery of the content and services. However, broadcasters don't yet have a detailed understanding of the delivered QoE in the same way they do for conventional broadcast networks; listeners meanwhile face the uncertainty of not knowing what level of expectation to have as to how well these services will work.

Project Timbre has begun to explore the role that real-world measurements and subsequent data analysis can have in addressing these issues, harnessing the inherent two-way nature of mobile IP networks to capture metrics directly from the test handsets in the project.

Some initial findings have been presented, including:

- Segment delivery time appears to be a useful proxy for QoE, that is independent of the length of the client buffer;
- The Time to First Byte is a significant element of the overall delivery time of each audio segment, and should be taken into account when estimating the effective bandwidth of the connection;
- Spatial and temporal aggregation of data reveals that QoE varies in both these dimensions, as well as from one network to another;
- The segment delivery time shows significant peak to mean variation. High QoE (of the type associated with broadcast network delivery) is associated with the tails of the associated distributions e.g. the 99$^{th}$ percentiles and above;
- It is possible to build-up a more complete picture of network performance, and to make comparisons with existing data sets, by the aggregation of measurements taken over an extended period of time; and
- Signal strength alone does not appear to reliably correlate with segment delivery time and, as such, there is a distinction between the ‘signal coverage’ and ‘service coverage’ offered by mobile networks.

More work is needed to identify the causes of, and to suggest solutions for, the issues set out in this paper. This is likely to be beyond both the resource and expertise of content providers alone. Similarly, mobile network operators do not have access to all the information required, including the performance of audio streaming services, themselves.

By using data to identify areas of concern, and to give better understanding of the real-world performance of mobile networks today, it is hoped to inform and encourage dialogue with the mobile industry – and other interested parties such as regulators – around how expectations of coverage for a given service can be better communicated to the user.

## FUTURE WORK

The work on Project Timbre is ongoing and a number of areas of further work have already been identified. The current analysis and results are solely based on the use of HTTP/1.1 over TCP for the delivery of audio segments over the network to the client. While limitations such as head of line blocking are likely to be less significant for the serial delivery of live audio segments, the alternative congestion algorithms available with HTTP/3 over QUIC are worthy of study as well as Low-Latency DASH techniques.

There is also the opportunity to continue to optimise and monitor future performance improvements based on further analysis of the dataset, such as the choice of initial client buffer size and its impact on service resilience.

Finally, there is a need to develop a better understanding of the role that techniques in the latest 3GPP standards such as the 5G Media Streaming (5GMS) System [5] could play in improving QoE for the listener.

5GMS, specified by 3GPP in TS 26.512 [6], allows Content Service Provider applications to actively collaborate with 5G networks to jointly achieve better outcomes for all users of the network. For example, differing Quality of Service envelopes to support real-time streaming or background downloads can be provisioned within the 5GMS system for a media session. These are then referenced by a Media Session Handler running as a background service on the handset.

In its simplest form, an app such as BBC Sounds could initiate media session handling by requesting just a 3GPP Service URL at the start of the streaming session. Meanwhile, a more sophisticated 5GMS-aware application could take advantage of network assistance during the media streaming session by, for example, receiving asynchronous recommendations of the bit rate that can currently be delivered reliably by the 5G network. This could be fed by the UE application into the media player's ABR algorithm to vary the streaming quality or by asking for a short bit rate ‘boost’ to allow the client to replenish its playback buffer following a short drop in network capacity or a loss of signal.

## REFERENCES


1. European Commission. Broadband Coverage in Europe 2022: Mapping progress towards the coverage. September 2023. Clause 4.3.4.3.
2. House of Commons Library Research Briefing. Rural mobile coverage in the UK: Not-spots and partial not-spots. March 2024. Clause 2.2.
3. UK. Shared Rural Network.
4. UK Government. Digital radio and audio review. April 2022.
5. Gabin, F., Lohmar, T., Heikkilä, G., D'Acunto, L. and Stockhammer, T. 5G Media Streaming Architecture. Proceedings of the IBC. October 2019.
6. 3GPP. 5G; 5G Media Streaming (5GMS); Protocols. TS 26.512.


## ACKNOWLEDGEMENTS


The authors would like to thank their colleagues across the BBC for their significant contributions to this work. They would also like to thank the International Broadcasting Convention for permission to publish this paper.