Survey of Mobile Handset Performance

Metrico Wireless, Ltd.
July 2011
INTRODUCTION

Metrico Wireless was commissioned by Ofcom to undertake a study of the performance, as perceived by mobile customers, of a range of mobile phones. This annex summarises findings for voice services.

The study was performed on 6 different models ranging from premium smart phone devices to relatively inexpensive entry level phones. The selected phones encompassed a range of different manufacturers and mobile phone operating systems.

Measurements were captured on key performance indicators that would be used by mobile subscribers to judge the quality of experience when using the devices. These included metrics for call completion and voice quality.

The studies were performed in two locations, London and Devon, representing a range of different coverage environments and also for alignment with previous Ofcom studies in Devon.

METHODOLOGY

2.1 OVERVIEW

A range of devices featuring different manufacturers, mobile operating systems and an array of retail price points were evaluated. The testing process, including the procurement of devices and subscriptions, was intended to be as close as possible to how the phones are used by consumers.

Metrico used its industry standard Nomad measurement solution to automate the capture of end to end voice quality and call retention performance for the mobiles under test.

2.1.1 Devices & Subscriptions

All devices were purchased over the counter, SIM free, from a range of vendors including Carphone Warehouse, Phones4U and Play.com. All subscriptions were purchased for the same network on rolling pay monthly contracts.

2.1.2 Reference Device

Metrico’s process quantifies the handset’s performance relative to a known reference phone (a 3G smartphone) enabling the results captured across different areas, days and devices to be summarised in the form of an overall scorecard describing relative device performance.

2.1.3 Drive Test Routes

The routes were designed to expose the mobiles to a range of coverage environments including:

- Good, moderate and marginal signal levels
- 3G and 2G plus 3G->2G network handovers
- Potentially coverage and interference limited network areas
Ofcom Survey of Mobile Handset Performance

The London loop covered an area including Hounslow, Harrow, Ealing, Kew and Twickenham and was primarily 3G coverage with small areas of 2G identified where the route is adjacent to the River Thames. It was anticipated that it would be primarily interference limited especially in exposed line of sight areas like Harrow on the Hill and Kew.

The Devon route covered the city centre of Plymouth and then a large loop covering Tavistock, the A386 running through Dartmoor National Park, Launceston, Callington and Saltash. It was anticipated that the route once outside Plymouth city boundaries would be 2G and coverage limited although there would be some 3G interference limited locations as the route crosses and runs adjacent to Plymouth Sound.

2.1.4 Handset mounting

The devices are mounted in suction cup car cradles on the interior windows of the vehicles with a minimum separation of 12 inches between phones. The phones are rotated around the cradles during the tests so each device spends roughly the same percentage of time in each position in the vehicle.

2.1.5 Email Background Traffic

In order to mimic real world customer usage models, email background traffic is sent to the smart phone devices. Small emails (approximately 1KB in size) are sent to the devices one every 60 seconds. The devices are set to retrieve email traffic at the highest update rate possible. The smart phones are receiving email traffic throughout the duration of all tests. Metrico’s collection system ensures that no more than 50 emails are resident on the device at any time, periodically purging older mails from the handset.

2.1.6 Engineering Mode phone

In order to characterise the routes and garner insight into the likely radio conditions the devices under test were experiencing, idle mode information was captured for each route using an engineering mode phone.

There will be variances in which radio bearer and levels the actual devices under test will be utilising given the difference in operation of handover and 3G re-selection between devices in network directed dedicated (in call) mode and idle mode.

2.1.7 Collection Dates/Times

Data was collected from April 4th to May 25th 2011 between the hours of 8am to 8pm Monday-Saturday.

2.2 CALL PERFORMANCE TESTING

In each location the call setup and retention performance of the devices was captured using the reference routes described in section 2.1.3. Each route was repeated numerous times over several
days until a minimum of 500 mobile originated call attempts had been placed on each device in each city.

The devices were tested in two runs, capturing data on 4 phones simultaneously on each run. The reference device was driven on both runs permitting normalisation of results across all phones.

An asynchronous call pattern (devices run the call pattern independent of one and other) was used to minimise the risk of generating localised network congestion. The call pattern consisted of a 120 second in call/60 second wait time, up to 30 second call setup time was allowed before the attempt was marked as a failed setup. The 60 second wait period following call completion or failure was designed to maximise the opportunity of 3G devices to reselect the 3G network.

### 2.3 Voice Quality Testing

In each location the end to end voice quality performance of the devices was captured using the reference routes described in section 2.1.3, the devices conducted a machine scored half duplex conversation using Metrico’s Nomad system until a minimum of 1000 (10 second) audio samples, typically representing 4-5 hours of recording, were captured for each device.

All calls are placed from the test vehicle to Metrico’s dedicated audio server, the call sequence used is an industry standard hold until drop pattern. If/when calls are dropped the engineer waits until service is re-established and then restarts the call. For 3G smart phones, if a device hands over to 2G, its call was terminated, a 60 second WAIT observed and then call re-initiated to give the device every opportunity to reselect 3G coverage. If the device re-initiates in 2G, process will be repeated after approximately 10 minutes of call time. The goal is to utilise 3G coverage whenever available (2G-3G in call handover not supported by current cellular standards). Given the routes are designed to expose the devices to a mix of 3G, 3G to 2G handover and 2G coverage, the voice quality results will include samples captured on 3G only, 2G only and 3G/2G calls.

Voice Quality results are expressed in terms of PESQ_LQ distributions and statistics. PESQ_LQ scores range from 1-4.5 and can be mapped to the listening scale below:

<table>
<thead>
<tr>
<th>Quality of Speech</th>
<th>PESQ_LQ/MOS Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;4</td>
<td>Excellent</td>
</tr>
<tr>
<td>4</td>
<td>Good</td>
</tr>
<tr>
<td>3</td>
<td>Fair</td>
</tr>
<tr>
<td>2</td>
<td>Poor</td>
</tr>
<tr>
<td>1</td>
<td>Bad</td>
</tr>
</tbody>
</table>

*Table 1: MOS Mapping*

Based on industry experience, Metrico characterizes PESQ/MOS scores based on the ranges below

<table>
<thead>
<tr>
<th>PESQ_LQ/MOS Range</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;= 3.6</td>
<td>Good/Acceptable Quality Voice Quality</td>
</tr>
<tr>
<td>&gt;=2.7 &amp; &lt; 3.6</td>
<td>Marginal Quality</td>
</tr>
<tr>
<td>&lt; 2.7</td>
<td>Poor/Unusable Quality</td>
</tr>
</tbody>
</table>

*Table 2: Thresholds for Voice Quality Scores*
3 RESULTS

3.1 CALL PERFORMANCE

This section contains the results for call setup and retention testing and outlines the performance of two device categories under test, multi-band smartphones and 2G/GSM capable feature/entry level phones. The difference in performance of the reference device (a multi-band smartphone) utilised on the two test runs is also presented in the sections below. The vertical bars represent the 90% confidence band around the result.

3.1.1 Overall Call Completion Success Rate

Call completion success is defined as the percentage of call attempts that are successfully placed on the network and retained for the full defined call duration.

London:

![Graph showing call completion success rates for London devices]

Devon:

![Graph showing call completion success rates for Devon devices]

Note: Charts show the mean values (the columns) and the lower and upper bounds of the 90th confidence bands (the vertical red lines).

Figure 1: Call Completion Success Rates
3.1.2 Dropped Call Rates

The dropped call rate is defined as the percentage of calls that are disconnected prior to the completion of the full defined call duration, divided by the number of call attempts that are successfully placed on the network.

London:

![Dropped Call Rates in London](chart)

Devon:

![Dropped Call Rates in Devon](chart)

Note: Charts show the mean values (the columns) and the lower and upper bounds of the 90th confidence bands (the vertical red lines).

Figure 2: Dropped Call Rates
3.1.3  Call Setup Success Rates

The call setup success rate is defined as the percentage of calls that are successfully placed on the network divided by the total number of call attempts.

London:

Devon:

Note: Charts show the mean values (the columns) and the lower and upper bounds of the 90th confidence bands (the vertical red lines).

Figure 3: Call Setup Success Rates

3.2  Voice Quality Performance

This section contains results for voice quality testing and outlines the performance of two device categories under test, multi-band smartphones and 2G/GSM capable feature/entry level phones. The difference in performance of the reference device (a multi-band smartphone) utilised on the two test runs is also presented in the sections below. When considering average MOS scores, a difference of 0.25 or more between the averages of two systems or devices can be perceived by the majority of listeners.
3.2.1 Voice Quality Summary

In London and Devon both handset categories returned MOS scores of over 3.6 on average to indicate good or acceptable voice quality for smart-phones and feature/entry level phones.

There were also no perceivable differences in the measured voice quality between handset categories, i.e. the difference in overall MOS between the smart-phones and feature/entry level phones was less than 0.25 in both London and Devon.

London:

Devon:

Note: The red horizontal bars at 3.60 MOS indicate that both categories of handsets returned on average good/acceptable voice quality results.

Figure 4: Average PESQ MOS (Mean Opinion Score)
4 OBSERVATIONS & CONCLUSIONS

The feature/entry level phones return better call setup results than the smartphones in both London and Devon, a statistically meaningful difference in Devon.

The feature/entry level phones return statistically better overall call completion rates in Devon relative to the smartphones. In London there was no statistically meaningful difference between the sets of devices.

It would appear from the data captured that the feature/entry level phones return better overall call setup and completion rates in areas of mixed 3G/2G and predominantly 2G network coverage represented in Devon.

Some of the factors that this behaviour may be attributable to are, the demands placed on smartphones simultaneously processing email and voice data, the increased intricacy of 3G/2G network re-selection & handover, along with the complexity and proliferation of antenna systems in smartphones (requirement for multiband support for 3G/2G, additional Wi-Fi hardware).

The feature/entry level phones recorded higher overall voice quality MOS in both London and Devon than the smartphones; however the difference in MOS between the two device categories was 0.07 in London and 0.11 in Devon. As a result, users would not perceive a difference in overall voice quality between the two sets of devices. (The majority of subscribers perceive a distinction in voice quality between two systems if average MOS scores differ by 0.25 MOS points or more.

The data collected in this study suggests that the MOS exhibited on smartphones decreases during 3G to 2G handover. This can have the effect of reducing the voice quality (MOS) during call as experienced by users.
5 ABOUT METRICO WIRELESS

The device market place is in the midst of a major upheaval resulting from a proliferation of new services and technologies, new device categories and entrants to the market. Subscribers increasingly cite device choice as a major factor in network preference and the device performance or perceived quality is a major factor in churn. Combinations of increasing competitive and pricing pressures accentuate the risk and financial consequence of launching a sub-standard device to market.

As poorly performing devices are not always reliably captured through conformance and regulatory testing, Metrico has introduced a ‘Fit4Launch’ and ‘Mobile Experience’ program for Network Operators and Device Manufacturers. This program ensures that devices perform to agreed standards in a range of real world and customer usage scenarios.

In partnership with the world’s leading wireless operators, handset manufacturers and wireless technology developers, Metrico helps the industry improve the mobile customer experience. Millions of wireless users are benefitting from the improved call performance, data services and voice quality. Metrico is the market leader in capturing the user experience and device quality across smart phones, tablets, data modems and Bluetooth accessories.

Using proprietary technology and methodologies, Metrico Wireless evaluates over 200 devices annually including 90% of all UMTS devices launched in the United States, with the goal of guaranteeing a positive end consumer experience. Our processes are independent of the underlying network technology and subscriber terminal. Working with operators worldwide, Metrico enables its customers to out-perform their competition by reducing handset returns and churn from poor user experience, yielding lower costs for new device introduction and accelerated time to market.