VoIP client benchmarking report

February 2007
Version 1.0

Prepared by:
Psytechnics Limited

23 Museum Street
Ipswich, Suffolk
United Kingdom
IP1 1HN

t : +44 (0) 1473 261 800
f : +44 (0) 1473 261 880
e : info@psytechnics.com

Contents

1. Executive Summary

Psytechnics have evaluated the speech quality performance of Microsoft's Office Communicator 2007 client using prototype USB handsets and Cisco 7961 IP phones with version 5.0 of the Cisco Call Manager. This document reports work conducted by Psytechnics on behalf of Microsoft.

The principal conclusions of the study were:

- Overall, the one-way listening speech quality provided by the combination of Microsoft's client and USB handset was consistently better than that provided by Cisco's IP phones and Call Manager, whether using G.711 or G.729.

- More detailed analysis showed that the mean transmission bandwidth of the Microsoft solution typically fell between 24kbps and 80kbps, which are the bit-rates of the Cisco solution with G.711 and G.729 operation respectively. The peak bandwidth of the Microsoft solution was 100kbps for some IP conditions.

- In general, the maximum recommended value for round-trip delay is approximately 400ms, i.e. a one-way delay of 200ms. The delays observed for both Clients with low IP impairment conditions are within, or close to, this budget, although it should be noted that control of the delay budget within VoIP networks remains essential.

Psytechnics is an established world leader in objective and subjective quality measurement and the provision of voice and video quality assessment solutions. A spin-off company from BT in 2000, Psytechnics have IPR in 6 ITU world standards, and an extensive subjective testing library which they continue to extend with new voice, audio and video subjective tests. Expertly designed subjective tests, such as those reported here, can be examined with well established statistical tests to validate independence of the tested variables and a sufficiency of conditions, subjects, test material, etc.

Psytechnics have supplied quality measurement software solutions to the communications industry worldwide. They have a state-of the-art multimedia subjective testing laboratory and regularly supply voice, video and audio subjective test data for international standardization of; objective measurement methods, CODECs and planning tools. This substantial subjective test database also provides excellent validation for Psytechnics' own objective quality measurement methods.

The study reported in this document was commissioned by Microsoft and conducted by Psytechnics as an independent third party expert. The report findings should not be taken as a purchasing recommendation from Psytechnics.

2. Introduction

The aim of the work described in this report was to measure the performance of the two VoIP clients under controlled test conditions representing a wide range of operational scenarios using a combination of subjective and objective experiments. These clients were a Beta release of Microsoft's Office Communicator 2007 with prototype USB handsets (Client 1), and a Cisco solution with 7961 IP Phones and version 5.0 of the Cisco Call Manager (Client 2). The experiments were based on speech transmitted between two VoIP clients of the same type. Each client was tested with two different voice codec algorithms enabled; thus four client/codec combinations were assessed in total.

The rest of this report is structured as follows:

- Section 3 provides information about the hardware and software configurations used in the benchmarking work.
- Section 4 describes the handset characterization that was performed in order to ensure that the experiments reflected true operating conditions.
- Section 5 provides a summary of the main experimental data.

- Section 6 provides the results of the subjective experiments and the associated statistical analysis; additional data from the experiments is provided in Annex A.
- Section 7 provides the results of objective analysis of the speech processing performed for the subjective experiments. The measurements made were transmission bit-rate and one-way delay.
- Section 8 contains a summary of the main conclusions drawn from the data.

3. VoIP client configurations

This section provides information about the hardware and software configurations used in the benchmarking work. Full details of the two VoIP clients are provided in Table 1 and Table 2.

Each IP impairment condition was processed with two different Client 1 codec configurations (default operation and narrowband operation) and two different Client 2 codec configurations (G.711 and G.729) as described in Table 1 and Table 2.

In its default mode of operation, Client 1 can switch between a narrowband codec and a wideband codec. In its default mode of operation Client 1 selected the wideband codec for all of the G.1050 IP conditions in the subjective experiment.

Table 1: VoIP Client 1 specification

Specification	**Description**
General description	PC-based VoIP client with external USB handset External call manager
USB Handset	Microsoft Catalina Prototype ES2 model 1106
VoIP client software	Microsoft Office Communicator 2007 (Beta 3)
PC Operating System	Windows XP Professional (Version 2002) SP2
PC hardware	Dell Optiplex GX520 3.0 GHz Intel Pentium 4 processor 512 Mbyte RAM
Call Manager Software	Office Communicator Server Standard Edition Beta 3 version
Call Manager Hardware	HP Proliant DL360 3.6 GHz Intel Xeon processor 2GByte RAM
Configuration	Silence suppression disabled
Codecs	Client 1 default codec selection behaviour Client 1 NB codec locked to narrowband codec

Table 2: VoIP Client 2 specification

Specification	Description
General description	IP Phone
Phone hardware	Cisco CP 7961G
IP Phone Firmware	Loadfile SCCP41.8-0-35 Boot Load ID boot41.3-2-2-0.bin OS Load ID cnu41.3-1-1-26.bin App Load ID jar41sccp.8-0-2-25.sbn DSP Load ID dsp41.1-1-2-26.sbn JVM Load ID cvm41sccp.8-0-2-25.sbn
Call Manager Software	System version: 5.0.4.2000-1 Administration version: 1.1.0.0-1
Call Manager Hardware	Cisco Media Convergence Server (MCS) 7825-H2
Configuration	Silence suppression disabled SCCP call control
Codecs	G.711 μ-law G.729

4. Handset characterization

In order to make the test calls as representative as possible, the speech was sent to and received from the VoIP clients using the analogue handset interfaces. It was therefore necessary to measure the send and receive filter characteristics of both handsets so that the test speech could be pre- and post-filtered to fully simulate the end-to-end connection. The pre- and post-filtering operations were performed using a 4096 point FFT filter approximation of the measured send and receive characteristics.

The handset characteristic measurements were made using a head and torso simulator (HATS) – see Figure 1 and Figure 2. Note that these photos are for illustration only, and do not show the handsets in their fully calibrated positions. Initial measurements were made using white noise to verify that third-octave measurements could accurately represent the transfer functions of the handset transducers. The send characteristics were then measured using a third-octave series of tones pre-normalized to produce 79dB SPL at the mouth reference point; the receive measurements were made using a third-octave series of tones normalized to 10dBm0 at the electrical input of the handset. The receive characteristic measurements were post-normalized to account for the transfer function between the ear and drum reference points of the HATS.

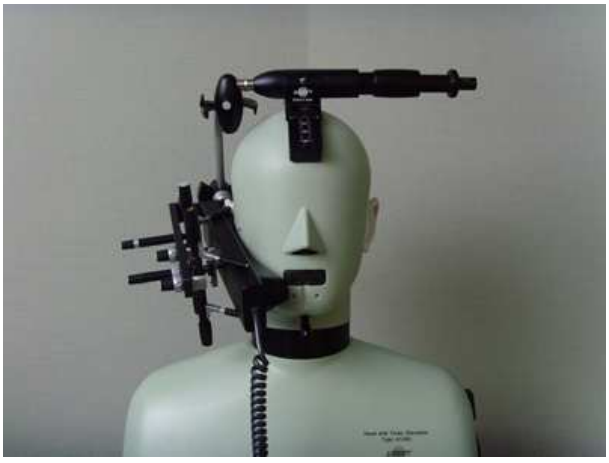

Figure 1: Client 1 Handset in HATS jig

Figure 2: Client 2 Handset in HATS jig

5. Summary of test conditions

This section provides a summary of the main experimental data for convenient reference. The subjective experiments were conducted according to ITU-T Rec. P.800 using the Absolute Category Rating (ACR) Listening Quality Scale, where subjects are asked to vote on each speech sample using the following 5-point voting scale: 1-Bad, 2-Poor, 3-Fair, 4-Good, 5-Excellent. Each experiment used speech samples from 6 different talkers (3 male and 3 female) and was assessed by 32 naive subjects. The basic design follows the balanced block design approach that has become the norm for all internationally co-ordinated ACR experiments run by the ITU-T. The experiment was performed twice: Experiment 1a was conducted using North American English speech material with no background noise (clean speech); Experiment 1b was conducted using British English speech material with office babble noise added at an SNR of 25dB.

5.1 Subjective test conditions

The condition list for the subjective experiments is provided in Table 3. Detailed descriptions of the IP impairment conditions are provided in Section 5.2.

Table 3: Condition list for subjective experiments

Condition	Processing	VoIP Client Codec	IP impairment	Audio bandwidth
1	16kHz sampled Direct			8kHz
2	MNRU Q=42dB			8kHz
3	MNRU Q=36dB			8kHz
4	MNRU Q=30dB			8kHz
5	MNRU Q=24dB			8kHz
6	MNRU Q=18dB			8kHz
7	MNRU Q=12dB			8kHz
8	G.722 at 56kbps			8kHz
9	G.722.1 at 24kbps			8kHz
10	8kHz sampled Direct			4kHz
11	G.711 μ-law at 64kbps			4kHz
12	G.729 at 8kbps			4kHz
13	VoIP Client 1	NB	IP 1	4kHz
14	VoIP Client 1	NB	IP 2	4kHz
15	VoIP Client 1	NB	IP 3	4kHz
16	VoIP Client 1	NB	IP 4	4kHz
17	VoIP Client 1	NB	IP 5	4kHz
18	VoIP Client 1	NB	IP 6	4kHz
19	VoIP Client 1	WB	IP 1	8kHz
20	VoIP Client 1	WB	IP 2	8kHz
21	VoIP Client 1	WB	IP 3	8kHz
22	VoIP Client 1	WB	IP 4	8kHz
23	VoIP Client 1	WB	IP 5	8kHz
24	VoIP Client 1	WB	IP 6	8kHz
25	VoIP Client 2	G.711/μ	IP 1	4kHz
26	VoIP Client 2	G.711/μ	IP 2	4kHz
27	VoIP Client 2	G.711/μ	IP 3	4kHz
28	VoIP Client 2	G.711/μ	IP 4	4kHz
29	VoIP Client 2	G.711/μ	IP 5	4kHz
30	VoIP Client 2	G.711/μ	IP 6	4kHz
31	VoIP Client 2	G.729	IP 1	4kHz
32	VoIP Client 2	G.729	IP 2	4kHz
33	VoIP Client 2	G.729	IP 3	4kHz
34	VoIP Client 2	G.729	IP 4	4kHz
35	VoIP Client 2	G.729	IP 5	4kHz
36	VoIP Client 2	G.729	IP 6	4kHz

5.2 IP impairment conditions

The IP impairment conditions used for the subjective experiments are summarized in Table 4 (zero burstiness should be interpreted as a uniform distribution).

The IP impairment conditions were derived from the ITU-T G.1050 model [1] because this model is able to simulate complex multi-link scenarios. The G.1050 settings corresponding to the these IP conditions are listed in Table 5 and Table 6, and the resultant packet loss and jitter measured for the conditions are reported in Table 7.

All conditions included a one-way 50ms bulk delay in addition to any jitter added by the IP models.

Table 4: Summary of IP impairment conditions for subjective experiments

Impairment	Scenario	Packet Loss	Burstiness	Jitter
IP 1	No IP impairments	-	-	-
IP 2	Good quality network	low	zero	Low
IP 3	Typical enterprise network	low	low	medium
IP 4	Multi-site connectivity	medium	high	High
IP 5	Internet conditions	high	medium	high
IP 6	Overloaded enterprise network	medium	high	medium/ high

Table 5: Fixed G.1050 parameters for subjective experiments

Parameter	Value
Core Parameters	
Base delay (ms)	50
Flap Delay (ms)	0
Flat Interval (sec)	3600
Link Fail Duration (ms)	0
Link Fail Interval (sec)	3600
LAN Parameters	
MTU Size (bytes)	1508
Duplex Mismatch (0/1)	0
MTU Size (bytes)	1508
Other Parameters	
Packet Size (bytes)	0

Table 6: G.1050 parameters varied as a function of IP impairment conditions for subjective experiments

Impairment	Jitter (ms)	Packet loss (%)	Out of seq. prob.	LAN speed (Mbps)	LAN occ. (%)	Link speed (kbps)	Link occ. (%)
IP 1	0	0	0.0	100	0	43000	0
IP 2	10	0	0.0	100	20	43000	0
IP 3	100	0	0.0	20	30	2048	10
IP 4	100	0	0.1	20	30	128/768	80
IP 5	400	15	0.1	20	40	128/768	80
IP 6	100	4	0.1	4	80	2048	20

Table 7: Long-term packet loss and jitter measured for G.1050 IP conditions

Impairment	Packet loss (%)	Min jitter (ms)	Max jitter (ms)	Mean abs jitter (ms)	Packet delay 90% percentile
IP1	0	-3	3	0.7	2
IP2	1.7	-12	14	3.5	11
IP3	3.4	-74	114	21.7	102
IP4	9.0	-143	220	27.2	348
IP5	25.4	-566	428	43.8	620
IP6	15.8	-295	493	36.9	365

6. Subjective experiments results

This section provides the results of the subjective experiments and the associated statistical analysis, and is divided into two parts:

- Sub-section 6.1 describes the results of Experiment 1a, which was conducted using North American English speech material with no background noise (clean speech);
- Sub-section 6.2 describes the results Experiment 1b, which was conducted using British English speech material with office babble noise added at an SNR of 25dB.

The purpose of the statistical analysis is to determine whether pairs of MOS scores can be considered to be significantly different at a given level of confidence. The analysis method used was Student's t-test, which is based on the mean and standard deviation of the subjective votes for the pair of MOS scores being compared. All comparisons were made at the 95% confidence level. It should be noted that the MOS values reported relate to the one-way listening quality and do not include the effects delay and echo, which can impact overall conversational quality.

Additional data from the experiments is provided in Annex A, including a breakdown of MOS by talker and 95% confidence intervals for all experimental conditions.

6.1 Experiment 1a: North American English, No Noise

6.1.1 Summary of results

The performance of the VoIP client configurations for different IP impairment conditions is shown in Figure 3. A comparison of the four VoIP client configurations with no IP impairments and the various reference codecs in the experiment is provided in Figure 4.

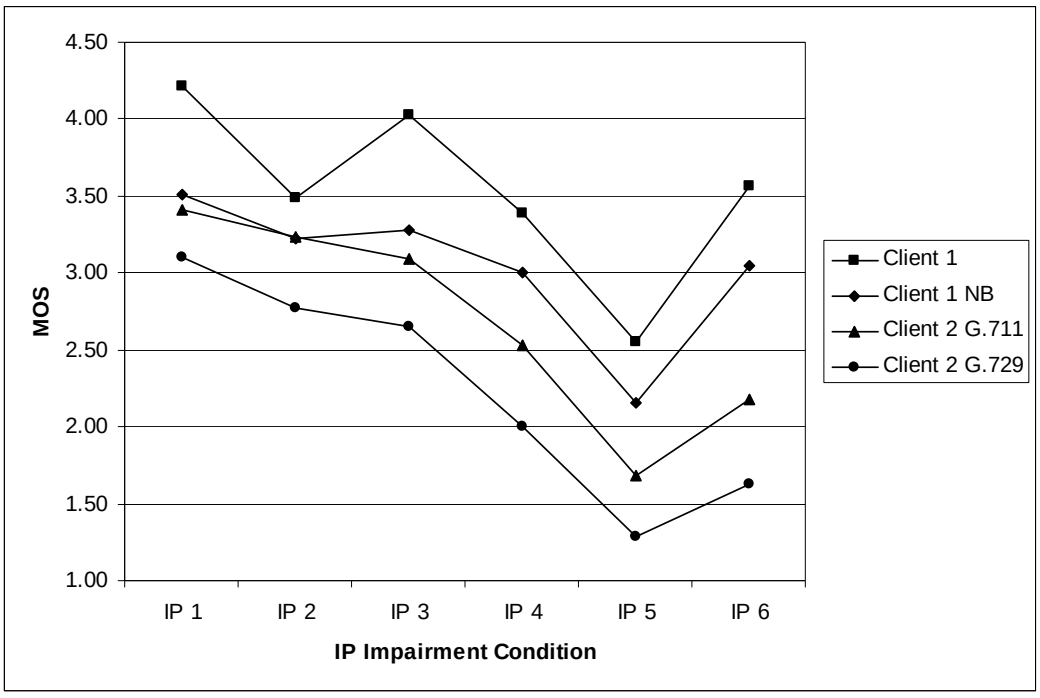

Figure 3: Performance of VoIP client configurations for different IP impairment conditions for Experiment 1a (North American English, no noise)

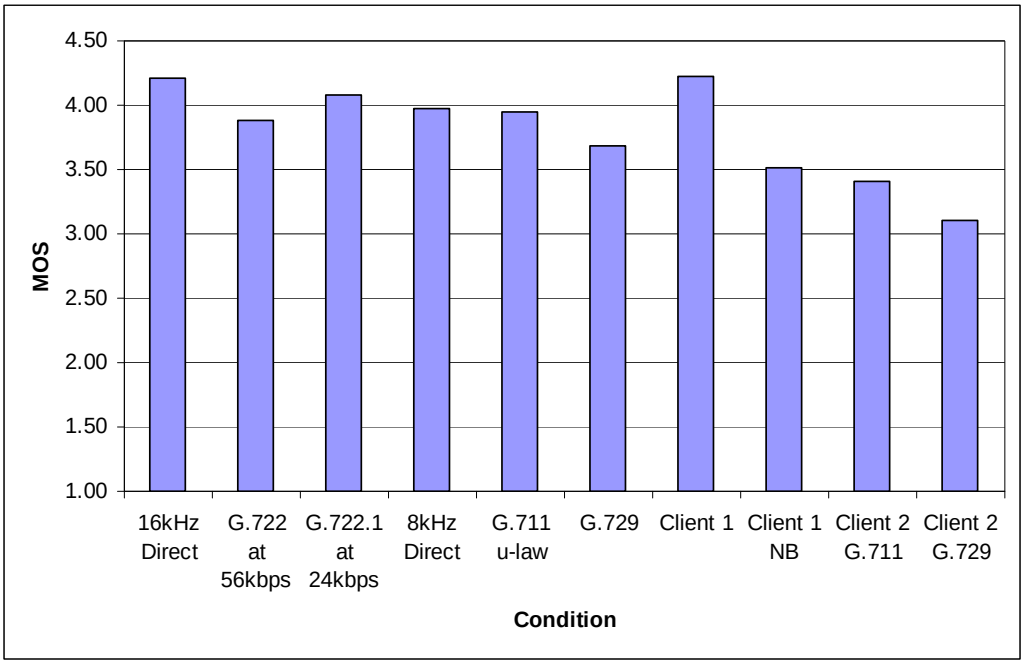

Figure 4: Comparison of VoIP client performance with no IP impairments with reference conditions for Experiment 1a (North American English, no noise)

6.1.2 Statistical analysis of subjective test data

The following sub-sections summarize the main results of a Student's t-test analysis of Experiment 1a.

6.1.2.1 Comparison of different VoIP client with IP impairments

For each IP impairment condition, the data was analyzed using Student's t-test at the 95% confidence level to establish whether the performance of the four VoIP client configurations were statistically different.

The results for Experiment 1a (North American English, no noise) may be summarized as follows:

1. Client 1 performed better than all other client configurations for all IP conditions.
2. When restricted to narrowband operation, Client 1 performed better than the G.729 mode of Client 2 for all IP conditions. It was also statistically equivalent to the G.711 mode of operation of Client 2 for IP conditions IP1 and IP2, and performed better for IP conditions IP3 to IP6.
3. The G.711 mode of Client 2 performed better than the G.729 mode operation for all IP conditions.

6.1.2.2 Comparison of VoIP clients with reference codecs

Table 8 shows the results of a t-test comparison at the 95% confidence level of the condition MOS value of the different client configurations with no IP impairments (IP1) and the various reference codecs included in the experiment. The contents of each cell should be interpreted as the relationship of the row to the column, i.e. the table shows that the Client 1 NB condition is worse than the 16kHz direct condition.

The results for Experiment 1a (North American English, no noise) may be summarized as follows:

1. The performance of Client 1 with no IP impairments was statistically equivalent to the unprocessed 16kHz direct condition and G.722.1 at 24kbps, and performed better than the G.722 at 56bkps, 8kHz direct, G.711 u-law and G.729 reference conditions.
2. When restricted to narrowband operation, Client 1 performed worse than all of the reference codec conditions.
3. Both the G.711 and G.729 modes of Client 2 performed worse than all of the reference codec conditions.

Table 8: Comparison of VoIP client configurations with reference codecs for Experiment 1a (North American English, no noise)

	16kHz Direct	G.722 at 56kbps	G.722.1 at 24kbps	8kHz Direct	G.711 u-law	G.729
Client 1	same	better	same	better	better	better
Client 1 NB	worse	worse	worse	worse	worse	worse
Client 2 G.711	worse	worse	worse	worse	worse	worse
Client 2 G.729	worse	worse	worse	worse	worse	worse

6.2 Experiment 1b: British English, Office Babble at 25dB SNR

6.2.1 Summary of results

The performance of the VoIP client configurations for different IP impairment conditions is shown in Figure 5. A comparison of the four VoIP client configurations with no IP impairments and the various reference codecs in the experiment is provided in Figure 6.

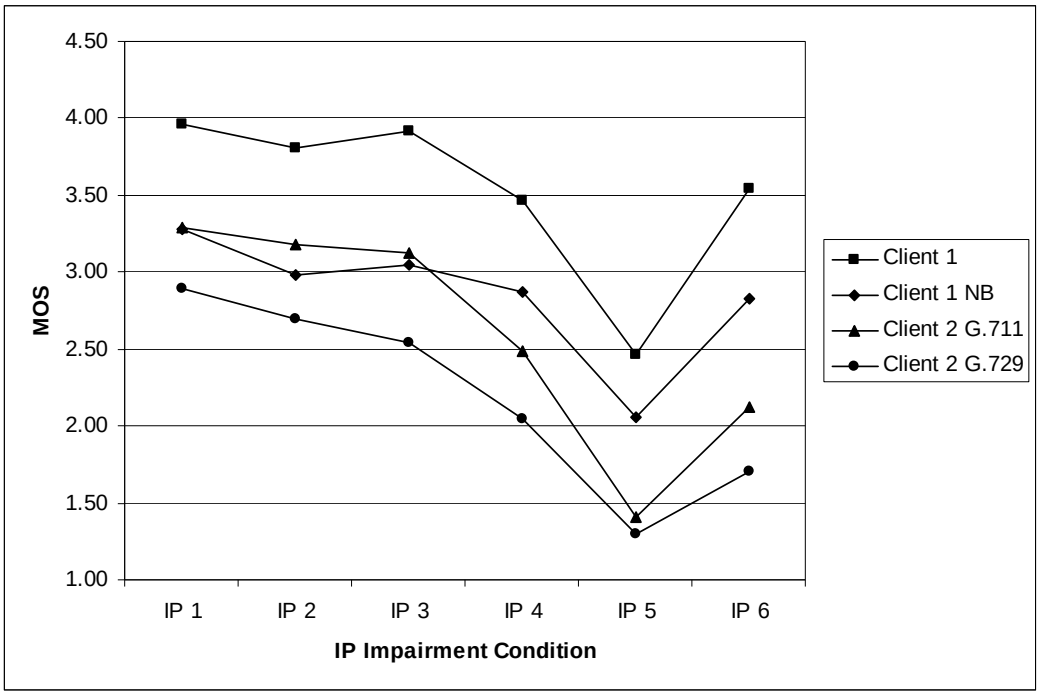

Figure 5: Performance of VoIP client configurations for different IP impairment conditions for Experiment 1b (British English, office babble at 25dB SNR)

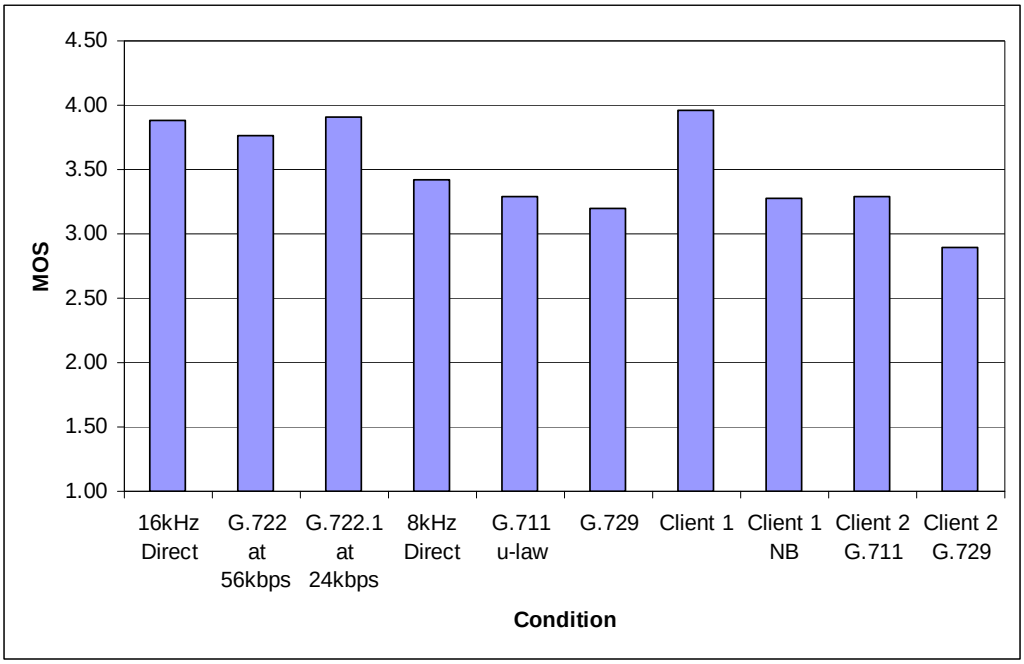

Figure 6: Comparison of VoIP client performance with no IP impairments with reference conditions for Experiment 1b (British English, office babble at 25dB SNR)

6.2.2 Statistical analysis of subjective test data

The following sub-sections summarize the main results of a Student's t-test analysis of Experiment 1b.

6.2.2.1 Comparison of different VoIP client with IP impairments

For each IP impairment condition, the data was analyzed using Student's t-test at the 95% confidence level to establish whether the performance of the four VoIP client configurations were statistically different.

The results for Experiment 1b (British English, office babble at 25dB SNR) may be summarized as follows:

1. Client 1 performed better than all other client configurations for all IP conditions.
2. The G.711 mode of Client 2 performed better than the G.729 mode operation of Client 2 for all IP conditions.
3. When Client 1 was restricted to narrowband operation and compared with the G.711 mode of Client 2, it was statistically equivalent for IP conditions IP1 and IP3, worse for IP condition IP2, and better for IP conditions IP4 to IP6.
4. When Client 1 was restricted to narrowband operation and compared with the G.729 mode of Client 2, it was better for all IP conditions.

6.2.2.2 Comparison of VoIP clients with reference codecs

Table 9 shows the results of a t-test comparison at the 95% confidence level of the condition MOS value of the different client conditions with no IP impairments (IP1) and the difference reference codecs included in the experiment. The contents of each cell should be interpreted as the relationship of the row to the column, i.e. the table shows that the Client 1 NB condition is worse than the 16kHz direct condition.

The results for Experiment 1b (British English, office babble at 25dB SNR) may be summarized as follows:

1. The performance of Client 1 with no IP impairments was statistically equivalent to the unprocessed 16kHz direct condition and G.722.1 at 24kbps, and performed better than the G.722 at 56bkps, 8kHz direct, G.711 u-law and G.729 reference conditions.
2. When restricted to narrowband operation, Client 1 was statistically equivalent to the 8kHz direct, G.711 u-law and G.729 conditions, and was worse than all of the wideband reference conditions.

3. The G.711 mode of Client 2 was statistically equivalent to the 8kHz direct, G.711 u-law and G.729 conditions, and was worse than all of the wideband reference conditions.
4. The G.729 mode of Client 2 performed worse than all of the reference codec conditions.

Table 9: Comparison of VoIP client configurations with reference codecs for Experiment 1b (British English, office babble at 25dB SNR)

	16kHz Direct	G.722 at 56kbps	G.722.1 at 24kbps	8kHz Direct	G.711 u-law	G.729
Client 1	same	better	same	better	better	better
Client 1 NB	worse	worse	worse	same	same	same
Client 2 G.711	worse	worse	worse	same	same	same
Client 2 G.729	worse	worse	worse	worse	worse	worse

7. Objective measurements

This section provides the results of objective analysis of the speech processing performed for the subjective experiments.

This section is divided into the following subsections, which correspond to the three types of measurement that were made for the different combinations of client configuration and IP condition:

- Subsection 7.1 reports bandwidth measurements;
- Subsection 7.2 reports one-way delay measurements.

7.1 Bandwidth

7.1.1 Measurement method

Packet capture traces were recorded for all VoIP client conditions processed. These were analyzed to determine the peak and mean transmission bit-rate as follows:

- The bit-rate calculations were restricted to the UDP port carrying the RTP media traffic, and do not include any signalling or RTCP packets.
- The calculations include the IP/UDP/RTP headers, but not any lower level headers, e.g. Ethernet.
- The mean bit-rate values are the transmission bandwidth, i.e. they do not include the effects of packet loss.
- The peak bit-rate was determined from the packet with the highest equivalent bit-rate, which is the number of bits in the packet divided by the duration of the packet in seconds.

7.1.2 Results

The mean transmission bit-rate for all combinations of client configuration and IP impairment condition are reported in Table 10. These values are also plotted in Figure 7.

7.1.3 Observations

The mean transmission bit-rate of Client 2 is independent of any IP impairment present, and is fixed at 80kbps for the G.711 mode of operation and 24kbps for the G.729 mode. The mean and peak bit-rates of Client 1 depend on the prevailing IP impairment as described below.

The Client 1 transmission bandwidth results may be summarized for the G.1050 IP impairment conditions as follows:

- The peak and mean bit-rates of Client 1 are always higher than the bit-rate of Client 2 in G.729 mode (24kbps).
- For IP impairments IP1 and IP2, the mean and peak bit rates of Client 1 WB (~50kbps and ~60kbps respectively) are lower than that of Client 2 in G.711 mode (~80kbps).
- For IP impairments IP3 to IP6, the mean bit-rate of Client 1 WB (80kbps) is approximately equal to that of Client 2 in G.711 mode, and the peak bit-rate of Client 1 (~100kbps) is higher.
- The peak and mean bit-rate of Client 1 NB is always lower than that of Client 2 in G.711 mode.

Table 10: Mean transmission bit-rate of clients as a function of IP condition (values are in kbps)

IP condition	Client 1 (default/WB)	Client 1 NB	Client 2 G.711	Client 2 G.729
IP 1	50.0	29.0	80.0	24.0
IP 2	50.0	30.2	80.0	24.0
IP 3	79.7	38.6	80.0	24.0
IP 4	79.1	38.6	80.0	24.0
IP 5	78.9	38.8	80.0	24.0
IP 6	79.6	39.4	80.0	24.0

Table 11: Peak transmission bit-rate of clients as a function of IP condition (values are in kbps)

IP condition	Client 1 (defdault/WB)	Client 1 NB	Client 2 G.711	Client 2 G.729
IP 1	60.0	37.6	80.0	24.0
IP 2	62.0	64.4	80.0	24.0
IP 3	99.2	66.4	80.0	24.0
IP 4	99.2	66.4	80.0	24.0
IP 5	99.2	65.6	80.0	24.0
IP 6	99.2	66.4	80.0	24.0

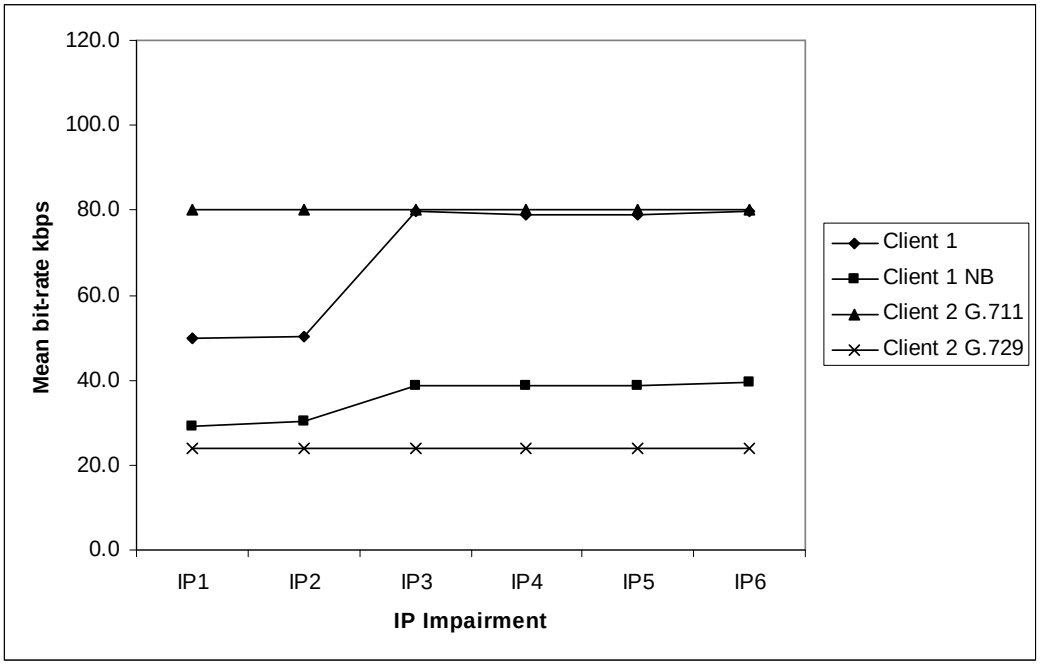

Figure 7: Mean transmission bit-rate of clients as a function of IP condition

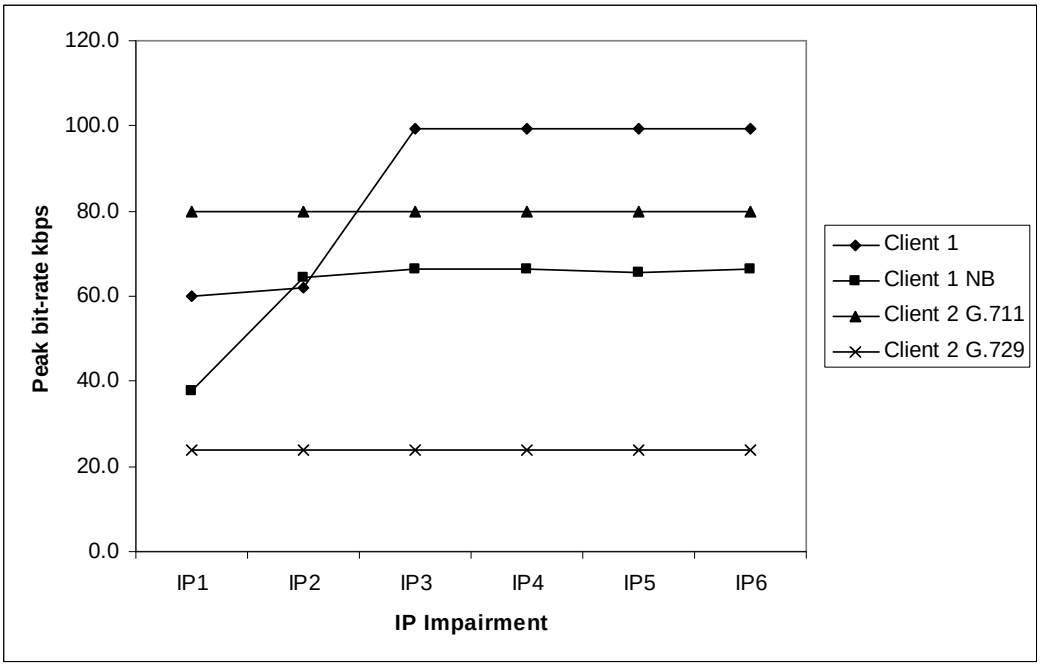

Figure 8: Peak transmission bit-rate of clients as a function of IP condition

7.2 Delay

7.2.1 Measurement method

The test set-up used to play and record speech material through the VoIP clients started the playback and recording processes simultaneously. This meant that it was possible to compare the sent and received file pairs to determine the one-way trip delay of the system. Since both clients were able to adapt the size of their jitter buffers to the prevailing IP conditions, it was necessary to use a sophisticated delay detection algorithm to measure the mean delay and a special tool developed by Psytechnics was used for this purpose.

7.2.2 Observations

In general, the maximum recommended value for round-trip delay is approximately 400ms, i.e. a one-way delay of 200ms. In high IP impairment conditions both clients can introduce a delay in excess of this recommendation.

8. Conclusions

This section contains a summary of the main conclusions drawn from the data.

8.1 Conclusions

The following are the main conclusions of the study (to aid clarity, comments relating to Client 1 have been limited to its default operation:

8.1.1 Speech Quality

The one-way listening speech quality results may be summarized as follows:

- The performance of Client 1 with no IP impairments was statistically equivalent to the unprocessed 16kHz direct condition in both subjective experiments.
- Client 1 performed better than both Client 2 configurations for all conditions in the subjective experiments.
- The G.711 mode of Client 2 performed better than the G.729 mode for all IP conditions in the subjective experiments.

8.1.2 Bandwidth

The mean transmission bit-rate of Client 2 is independent of any IP impairment present, and is fixed at 80kbps for the G.711 mode of operation and 24kbps for the G.729 mode of operation. The mean and peak bit-rates of Client 1 depend on the prevailing IP impairment as described below.

The Client 1 transmission bandwidth results may be summarized as follows:

- The peak and mean bit-rates of Client 1 are always higher than the bit-rate of Client 2 in G.729 mode (24kbps).
- For IP impairments IP1 and IP2, the mean and peak bit rates of Client 1 (~50kbps and ~60kbps respectively) are lower than that of Client 2 in G.711 mode (80kbps).
- For IP impairments IP3 to IP6, the mean bit-rate of Client 1 (~80kbps) is approximately equal to that of Client 2 in G.711 mode, and the peak bit-rate of Client 1 (~100kbps) is higher.

8.1.3 Delay

The delay results may be summarized as follows (all references to delay are one-way values; all conditions included 50ms of one-way bulk delay):

- In general, the maximum recommended value for round-trip delay is approximately 400ms, i.e. a one-way delay of 200ms. The delays observed for both Clients with low IP impairment conditions are within, or close to, this budget.

9. References

[1] "Network model for evaluating multimedia transmission performance over internet protocol", ITU-T Recommendation G.1050 (11/2005).

10. Abbreviations

CI	Confidence Interval
ERP	Ear Reference Point
HATS	Head and Torso Simulator
MRP	Mouth Reference Point
NB	Narrowband
SNR	Signal-to-Noise Ratio
SPL	Sound Pressure Level
WB	Wideband

Annex A: Additional subjective test data

Table 12 and Table 13 contain the condition MOS values, standard deviations and 95% confidence intervals for Experiment 1a and Experiment 1b respectively.

The 95% confidence interval indicates the range around the sample mean (± the CI value) within which there is a probability of 0.95 that the true mean lies. Note that the Student's t-test results reported in the main body of the report are calculated directly from the mean and standard deviation of the subjective votes for a pair of conditions, not by comparing the CI values of the conditions.

Table 12: MOS and Confidence intervals for Experiment 1a (North American English, Clean Speech)

Condition	Processing	VoIP Client Codec	IP impairment	MOS	stdev	95% conf. interval
1	16kHz sampled Direct			4.21	0.72	0.103
2	MNRU Q=42dB			3.91	0.75	0.107
3	MNRU Q=36dB			3.44	0.83	0.118
4	MNRU Q=30dB			2.74	0.77	0.110
5	MNRU Q=24dB			2.16	0.68	0.097
6	MNRU Q=18dB			1.69	0.70	0.100
7	MNRU Q=12dB			1.26	0.49	0.070
8	G.722 at 56kbps			3.88	0.87	0.124
9	G.722.1 at 24kbps			4.08	0.72	0.103
10	8kHz sampled Direct			3.97	0.77	0.109
11	G.711 μ-law at 64kbps			3.95	0.77	0.110
12	G.729 at 8kbps			3.69	0.71	0.101
13	VoIP Client 1	NB	IP 1	3.51	0.78	0.111
14	VoIP Client 1	NB	IP 2	3.22	0.76	0.108
15	VoIP Client 1	NB	IP 3	3.28	0.78	0.111
16	VoIP Client 1	NB	IP 4	3.00	0.85	0.121
17	VoIP Client 1	NB	IP 5	2.15	0.91	0.130
18	VoIP Client 1	NB	IP 6	3.05	0.78	0.112
19	VoIP Client 1	WB	IP 1	4.22	0.70	0.101
20	VoIP Client 1	WB	IP 2	3.49	0.88	0.126
21	VoIP Client 1	WB	IP 3	4.03	0.77	0.110
22	VoIP Client 1	WB	IP 4	3.39	0.89	0.127
23	VoIP Client 1	WB	IP 5	2.55	1.09	0.156
24	VoIP Client 1	WB	IP 6	3.57	0.77	0.110
25	VoIP Client 2	G.711/μ	IP 1	3.41	0.83	0.118
26	VoIP Client 2	G.711/μ	IP 2	3.23	0.77	0.110
27	VoIP Client 2	G.711/μ	IP 3	3.09	0.67	0.096
28	VoIP Client 2	G.711/μ	IP 4	2.53	0.82	0.117
29	VoIP Client 2	G.711/μ	IP 5	1.68	0.76	0.108
30	VoIP Client 2	G.711/μ	IP 6	2.18	0.81	0.115
31	VoIP Client 2	G.729	IP 1	3.10	0.72	0.102
32	VoIP Client 2	G.729	IP 2	2.78	0.81	0.116
33	VoIP Client 2	G.729	IP 3	2.65	0.78	0.111
34	VoIP Client 2	G.729	IP 4	2.01	0.76	0.108
35	VoIP Client 2	G.729	IP 5	1.29	0.55	0.078
36	VoIP Client 2	G.729	IP 6	1.63	0.72	0.103

Table 13: MOS and Confidence intervals for Experiment 1b (British English, Office Babble at 25dB SNR)

Condition	Processing	VoIP Client Codec	IP impairment	MOS	stdev	95% conf. interval
1	16kHz sampled Direct			3.88	0.86	0.122
2	MNRU Q=42dB			3.86	0.85	0.121
3	MNRU Q=36dB			3.74	0.86	0.122
4	MNRU Q=30dB			3.19	0.87	0.124
5	MNRU Q=24dB			2.69	0.80	0.114
6	MNRU Q=18dB			2.17	0.71	0.102
7	MNRU Q=12dB			1.63	0.63	0.091
8	G.722 at 56kbps			3.76	0.91	0.130
9	G.722.1 at 24kbps			3.91	0.79	0.113
10	8kHz sampled Direct			3.42	0.75	0.107
11	G.711 μ-law at 64kbps			3.29	0.73	0.104
12	G.729 at 8kbps			3.19	0.79	0.113
13	VoIP Client 1	NB	IP 1	3.28	0.79	0.112
14	VoIP Client 1	NB	IP 2	2.98	0.79	0.113
15	VoIP Client 1	NB	IP 3	3.05	0.77	0.110
16	VoIP Client 1	NB	IP 4	2.87	0.84	0.119
17	VoIP Client 1	NB	IP 5	2.06	0.95	0.136
18	VoIP Client 1	NB	IP 6	2.83	0.73	0.104
19	VoIP Client 1	WB	IP 1	3.96	0.73	0.105
20	VoIP Client 1	WB	IP 2	3.80	0.75	0.107
21	VoIP Client 1	WB	IP 3	3.91	0.78	0.111
22	VoIP Client 1	WB	IP 4	3.47	1.00	0.143
23	VoIP Client 1	WB	IP 5	2.46	1.09	0.156
24	VoIP Client 1	WB	IP 6	3.54	0.91	0.130
25	VoIP Client 2	G.711/μ	IP 1	3.29	0.80	0.114
26	VoIP Client 2	G.711/μ	IP 2	3.18	0.76	0.108
27	VoIP Client 2	G.711/μ	IP 3	3.13	0.75	0.107
28	VoIP Client 2	G.711/μ	IP 4	2.49	0.81	0.116
29	VoIP Client 2	G.711/μ	IP 5	1.41	0.58	0.083
30	VoIP Client 2	G.711/μ	IP 6	2.12	0.79	0.112
31	VoIP Client 2	G.729	IP 1	2.89	0.79	0.113
32	VoIP Client 2	G.729	IP 2	2.70	0.75	0.107
33	VoIP Client 2	G.729	IP 3	2.54	0.77	0.110
34	VoIP Client 2	G.729	IP 4	2.04	0.79	0.113
35	VoIP Client 2	G.729	IP 5	1.30	0.60	0.085
36	VoIP Client 2	G.729	IP 6	1.70	0.76	0.108