Implementer's Guide I for Recommendation ITU-T P.863: Mapping function of P.863 results into MOS-LQO

1 Introduction & background

Recommendation ITU-T P.863 was approved in early February 2011. During the last SG12 meeting it was discussed and agreed to develop a mapping function from ITU-T P.863,scores into MOS-LQO. This mapping function should be a compromise for linearization of the scores for a wide set of experiments it should meet the scale boundaries as set for P.863 and it should avoid the systematic offset described in [COM12-C159, 2010].

The authors have developed a mapping function based on all experiments used in the selection phase of P.863 and have validated it on individual experiments not used in this phase.

Finally, the authors decided on two functions, one for the P.863 narrowband operational mode and one for the P.863 super-wideband mode. Both functions are 3rd-order polynomials that are monotonic within the considered range of the scale.

It was further decided to use these functions 'on top' of the existing scale fitting in P.863. This requires no change to the P.863 model as described in the Recommendation.

2 Mapping functions

The mapping function is given by the following formula, where *MOSLQO* is the mapped value and *PMOS* the prediction made by P.863:

$$MOSLQO = a_0 + a_1\, PMOS + a_2\, PMOS^2 + a_3\, PMOS^3$$

With the following coefficients:

	a_0	a_1	a_2	a_3
narrowband	0.79	0.0036	0.2117	-0.0065
super-wideband	0.276	0.7203	-0.00756	0.01141

The shape of the polynomials is depicted by Figure 1.

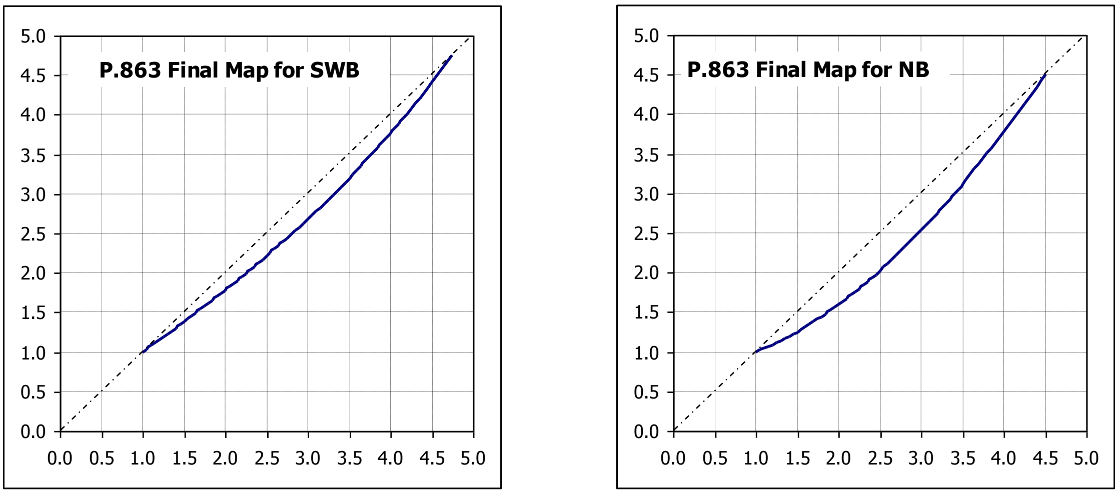

Figure1 - Shape of the mapping functions used for NB and SWB mode of P.863

3 P.863 performance results with new mapping functions

This section shows the results of the statistical evaluation of P.863 results with the new mapping functions. In order to allow for a direct comparison, the same statistical methods as in the selection phase were used. For comparison the 'unmapped' values are given in a second column.

In addition to the rmse* after 3rd- and 1st-order fitting, the rmse* without any fitting (rmse* raw) is shown as well.

Database	*rmse* 3rd*	*rmse* 1st*	*rmse* 3rd*	*rmse* 1st*	*rmse* raw*
Set 1	**P.863**	**P.863**	**P.863+Map**	**P.863 + Map**	**P.863 + Map**
NB_BT_P862_BGN_ENG	0.0981	0.1269	0.0988	0.1933	0.3243
NB_BT_P862_PROP	0.1658	0.1684	0.1665	0.1682	0.4458
NB_DT_P862_1st	0.1473	0.1830	0.1488	0.1659	0.3689
NB_DT_P862_BGN_GER	0.1112	0.1132	0.1114	0.1549	0.1846
NB_DT_P862_Share	0.0895	0.0893	0.0897	0.0904	0.1848
NB_ERIC_AMR_4B	0.1356	0.1573	0.1339	0.1401	0.1598
NB_ERIC_P862_NW_MEAS	0.1767	0.1759	0.1754	0.1769	0.2822
NB_TNO_P862_KPN_KIT97	0.1891	0.2089	0.1866	0.2100	0.2540
NB_TNO_P862_NW_EMU	0.1530	0.1567	0.1528	0.1529	0.4758
NB_TNO_P862_NW_MEAS	0.1654	0.1724	0.1650	0.1841	0.3567
NB_ITU_SUPPL23_EXP1a	0.1184	0.1195	0.1134	0.1223	0.1394
NB_ITU_SUPPL23_EXP1d	0.0676	0.0661	0.0675	0.0677	0.3978
NB_ITU_SUPPL23_EXP1o	0.1096	0.1232	0.1025	0.1122	0.1214
NB_ITU_SUPPL23_EXP3a	0.1660	0.2035	0.1686	0.2211	0.5095
NB_ITU_SUPPL23_EXP3c	0.0862	0.0989	0.0860	0.1341	0.2339
NB_ITU_SUPPL23_EXP3d	0.0585	0.0646	0.0581	0.0707	0.2949
NB_ITU_SUPPL23_EXP3o	0.0569	0.0623	0.0562	0.0841	0.0879
NB_FT_P563_PROP	0.0662	0.0653	0.0663	0.0701	0.0765
NB_LUC_P563_PROP	0.0926	0.1234	0.0917	0.1141	0.1323
NB_OPT_P563_PROP	0.1198	0.1179	0.1189	0.1265	0.1896
NB_PSY_P563_PROP	0.1736	0.1848	0.1751	0.2053	0.2082
NB_SQ_P563_PROP	0.1701	0.1815	0.1706	0.1732	0.5917
Average	**0.1235**	**0.1347**	**0.1229**	**0.1426**	**0.2736**

Database	*rmse* 3rd*	*rmse* 1st*	*rmse* 3rd*	*rmse* 1st*	*rmse* raw*
Set 2	**P.863**	**P.863**	**P.863+Map**	**P.863 + Map**	**P.863 + Map**
NB_ATT_iLBC	0.1937	0.2305	0.1945	0.2149	0.3244
NB_ERIC_Field_GSM_EU	0.1546	0.1548	0.1550	0.1686	0.2953
NB_ERIC_Field_GSM_US	0.1454	0.1491	0.1458	0.1477	0.1904
NB_GIPS_EXP1	0.1019	0.1257	0.1023	0.1677	0.1745
WB_GIPS_EXP3	0.1341	0.1530	0.1379	0.1808	0.1904
SWB_GIPS_EXP4	0.0777	0.0791	0.0781	0.0774	0.3212
NB_QUALCOMM_EXP1b	0.1206	0.1232	0.1110	0.1287	0.2312
WB_QUALCOMM_EXP1w	0.1269	0.1384	0.1217	0.1513	0.1792
NB_QUALCOMM_EXP2b	0.1491	0.1489	0.1446	0.1452	0.2601
NB_QUALCOMM_EXP3w	0.0956	0.0972	0.0883	0.0979	0.1859
WB_QUALCOMM_EXP3w	0.0708	0.1099	0.0727	0.1494	0.2009
SWB_48kHz101_ERICSSON	0.2804	0.2871	0.2811	0.2989	0.4086
WB_48kHz102_ERICSSON	0.1936	0.1920	0.1939	0.1972	0.2606
SWB_48kHz201_FT_DT	0.2717	0.2953	0.2741	0.2771	0.3966
SWB_48kHz202_FT_DT	0.2499	0.2446	0.2490	0.2542	0.3098
SWB_48kHz301_OPTICOM	0.2773	0.2835	0.2760	0.3024	0.2982
SWB_48kHz302_OPTICOM	0.1933	0.2004	0.1934	0.2182	0.3162
SWB_48kHz401_PSYTECHNICS	0.1474	0.1485	0.1474	0.1682	0.1868
WB_16kHz402_PSYTECHNICS	0.1839	0.1838	0.184	0.1800	0.3785
SWB_48kHz501_SWISSQUAL	0.1712	0.1915	0.1714	0.2087	0.2097
SWB_48kHz502_SWISSQUAL	0.2637	0.2592	0.2635	0.2650	0.3298
SWB_48kHz601_TNO	0.2199	0.2186	0.2201	0.2187	0.3219
SWB_48kHz602_TNO	0.1692	0.1887	0.1720	0.1972	0.2083
	0.1738	**0.1825**	**0.1730**	**0.1917**	**0.2661**

Database	*rmse* 3rd*	*rmse* 1st*	*rmse* 3rd*	*rmse* 1st*	*rmse* raw*
Set 3	**P.863**	**P.863**	**P.863+Map**	**P.863 + Map**	**P.863 + Map**
SWB_48kHz103_ERICSSON	0.2203	0.2275	0.2209	0.2412	0.3276
NB_8kHz104_ERICSSON	0.2840	0.2788	0.2841	0.2820	0.4369
SWB_48kHz203_FT_DT	0.2812	0.2883	0.2811	0.2793	0.2902
WB_16kHz204_FT_DT	0.2319	0.2325	0.2298	0.2301	0.2730
SWB_48kHz303_OPTICOM	0.1785	0.1805	0.1794	0.1982	0.3024
SWB_48kHz403_PSYTECHNICS	0.1697	0.1671	0.1697	0.1718	0.1837
NB_48kHz404_PSYTECHNICS	0.1614	0.1661	0.1596	0.1589	0.5422
SWB_48kHz503_SWISSQUAL	0.1946	0.1950	0.1957	0.1948	0.2007
NB_8kHz504_SWISSQUAL	0.2311	0.2355	0.2310	0.2314	0.4076
SWB_48kHz603_TNO	0.1602	0.1570	0.1603	0.1660	0.1665
NB_8kHz_NTT_PTEST_1	0.0872	0.0916	0.0877	0.0878	0.6092
NB_QUALCOMM_EXP4	0.1254	0.1265	0.1226	0.1339	0.3962
WB_QUALCOMM_EXP5	0.1100	0.1413	0.1111	0.1718	0.1890
NB_QUALCOMM_EXP6a	0.2164	0.2130	0.2163	0.2262	0.3177
NB_QUALCOMM_EXP6b	0.1191	0.1373	0.1183	0.1823	0.3151
NB_16kHz_HUAWEI_1	0.1317	0.1288	0.1271	0.1319	0.3172
NB_16kHz_HUAWEI_2	0.2015	0.2283	0.1922	0.2331	0.3439
	0.1826	**0.1879**	**0.1816**	**0.1953**	**0.3305**

Abs. worst case performance over the three sets				
Absolute worst case	**0.2840**	**0.2953**	**0.2841**	**0.3024**
Avg. Of three worst experiments	**0.2819**	**0.2847**	**0.2821**	**0.2867**

Contribution [COM12-C159, 2010] revealed a systematic positive prediction offset across all data sets of >0.2 MOS. The average prediction error was defined as

$$AvgPerror = \text{sum}\ (MOSLQO - MOSLQS)\ /\ N,$$

where $MOSLQS$ is the subjective MOS per condition and $MOSLQO$ the P.863 score and N the number of all conditions considered.

In case the new mapping functions are applied, the average prediction error AvgPerror is 0.041 for super-wideband and 0.081 for the narrowband operational mode, respectively. These values show that there is almost no systematic positive offset anymore as described in [COM12-C159] where this value exceeded 0.2.

4 Discussion

The mapping functions are a compromise between linear behaviour to subjective MOS over a wide spread of databases and the desired scale boundaries of 1.0 to 4.5 in narrowband mode and 1.0 to 4.75 in super-wideband mode. These constraints result in a slight non-linearity of the mapped MOS at the very high scale boundary. However, the mapping results show an average rmse* 1^{st} and 3^{rd} statistically insignificant different from the original values obtained for P.863.

For the determining the mapping functions no databases were specifically excluded since such exclusions are always subject to individual judgement. For the narrowband case all narrowband databases of the P.863 selection phase were used. For the super-wideband case all super-wideband databases were considered. The traditional wideband databases were however not used for deriving the super-wideband mapping in order to avoid any kind of bias.

The mapping was made based on per-condition results to avoid an over-weighting of experiments conducted in partially or even fully fractional design, where the number of files per condition is disproportionally higher than for the majority of non-fractional designed experiments.

5 Additional Code for P.863 Mapping

New Code in Time.cpp inserted starting from Line 512 in Public C Code:

```
if (aListeningCondition == STANDARD_IRS)
{
	predictedMosOverall = 0.79 + 0.0036*predictedMosOverall + 0.2117*predictedMosOverall*predictedMosOverall - 0.0065*pow(predictedMosOverall,3);

	if (predictedMosOverall < 1.00) predictedMosOverall = 1.00;
	if (predictedMosOverall > 4.50) predictedMosOverall = 4.50;
}

if (aListeningCondition == WIDE_H)
{
	predictedMosOverall = 0.276 + 0.7203*predictedMosOverall - 0.00756*predictedMosOverall*predictedMosOverall + 0.01141*pow(predictedMosOverall,3);

	if (predictedMosOverall < 1.00) predictedMosOverall = 1.00;
	if (predictedMosOverall > 4.75) predictedMosOverall = 4.75;
}
```

NB Mode (limitation to [1;4.5]):

MOSLQO = 0.79 + 0.0036*POLQAScore + 0.2117*POLQAScore^2 – 0.0065* POLQAScore^3

SWB Mode (limitation to [1;4.75]):

MOSLQO = 0.276 + 0.7203*POLQAScore - 0.00756* POLQAScore^2 + 0.01141* POLQAScore^3

____