Tdoc S4-110763

Source: Audience, Inc.

Title: Correlation of ETSI EG 202 396-3 Objective Speech Quality Measures to P.835 Subjective Test Results

Document for: Discussion

Agenda Item: 9

1. Introduction

Currently, the most accurate method for evaluating quality of noise reduced signals is subjective tests. Conversational tests are the unique way to naturally take into account all effects due to background noise, such as Lombard effect for example. Listening tests according to ITU-T Recommendation P.835 [1] however represent another reliable and much less expensive (time, money) solution: ITU-T P.835 was defined for the difficult task of subjective evaluation of noise reduced signals, it was proposed to solve the bi-dimensional problem linked to the evaluation of noise reduced signals (some users focus on the background noise whereas others on the distortions produced on the speech signal). The protocol consists in asking the listener to successively attend to and rate the processed speech signal on: i) the speech signal distortion alone using a five-point scale; ii) the background noise intrusiveness alone using a five-point scale; iii) the overall quality using the classical ACR (Absolute Category Rating) MOS scale. The principle of P.835 is to draw listener's attention on both types of distortion which may be present on noise reduced signals. This avoids the natural human behavior which would focus on only one type of degradation without being aware of the presence of the other one.

The currently in-force version of P.835 includes Amendment I Appendix III [4], which specifies a range of signal-to-noise ratios (SNR = 0, 6 and 12dB) and a variety of 6 different noise types (street, car, babble, pink, single-voice, and music).

There is only one objective measurement method available to assess the quality of noise reduced signals in a similar way as P.835: ETSI EG 202 396-3 (11/2008) [2]. ETSI EG 202 396-3 provides a hearing based prediction model allowing to predict the speech-, noise- and overall-quality in background noise situations as perceived subjectively by the user. In this study, we tested the commercial implementation of ETSI EG 202 396-3 offered by Head Acoustics as 3Quest.

In the following we present and discuss the result of tests performed by Audience on narrow-band mobile devices with a two-microphone noise suppressor to analyze the correlation between listening tests according to ITU-T P.835 and objective measures according to ETSI EG 202 396-3. The goal is to determine whether ETSI EG 202 396-3 would be suitable for estimating the performance of two-microphone noise reduction features embedded in mobile phones in a variety of real background noise conditions. This study follows a similar study by Orange described in TSG-SA4 #62 S4 (11) 0085 and TSG-SA4 #63 S4-110277.

2. Test Setup

The Noise Suppressor under test was a two-microphone noise suppressor using a multiplicative spectral energy suppression strategy (i.e. pure suppression, no cancellation).

Audio data and Subjective Listening Scores were collected for a broad-ranging sweep of input Signal-to-Noise Ratios (SNR) and Noise Suppression (NS) strength, with SNRs ranging from 0 to 30 dB in 6 dB steps (6 conditions), and NS values ranging from 0 to 30 dB in 3 dB steps, and 35 dB (12 conditions), for a total of 6 x 12 = 72 conditions. The SNR is determined at the primary microphone of the device. Noise Suppression was held constant throughout a given test, and no post-processors (AGC, Post Equalizers) were active. A P.835 Amendment I Appendix III [4] methodology was followed, with a listener panel of 32 naïve listeners. The noise source for the test was the “pub” sample from the ETSI EG 202 396-1 database. The speech samples for this investigation were recordings based on two different Talker Sets:

Talker Set 1	Talker Set 2
Speech samples from the P.501 database, with 4 talkers (2 male, 2 female) with two sentences each, for a total of eight sentences for each measurement. The files for all were concatenated into a single file containing all eight sentences, and a single objective ETSI EG 202-396-3 measurement was made for the concatenated file, as in the recommended procedure for 3Quest. Used in Primary Experiment reported in Figures 1 and 2.	Speech samples from Audience’s cell-phone database, with 4 talkers (2 male, 2 female) with one sentence each, each sentence repeated once, for a total of eight sentences for each measurement. Each talker’s file was evaluated separately by ETSI EG 202-396-3, and the four scores were averaged to produce a final score for each condition. Used in Talker Dependence Experiment reported in Figure 4.

The speech samples were convolved with measured impulse responses from mouth simulator to primary and secondary microphones of a reference handset mounted on a Head Acoustics Head and Torso Simulator (HMS II-3). The noise samples were convolved with measured impulse responses from the speakers in a ETSI EG 202 396-1 4-loudspeaker setup to the primary and secondary mics, with the left channel played through the two left speakers, and the right channel played through the two right speakers. The distance between the center of the test arrangement and the loudspeakers was 1.5 meters.

The speech sample levels were set at -26dBoV at the primary microphone, and the noise levels were set with A-weighting so as to produce SNRs ranging from 0 to 30 dB in 6 dB steps. No additional filtering of input reference files was done; 3Quest applies an IRS SND filter to those files by default. IRS RCV filtering of the noise-suppressed processed files was enabled, as in the recommended procedure for 3Quest. We did not enable variable delay detection, since this was believed to be intended to handle data from real networks in which packet loss may have occurred, a condition which does not occur in our experiment.

The calculation of S-MOS, N-MOS and G-MOS was made according to ETSI EG 202 396-3 using the commercial implementation of 3Quest from Head Acoustics.

A subjective test according to ITU-T P.835 was also performed on the recorded samples, with each trial in the subjective test coming from a single sentence from the speech samples. Each of the 32 subjects in the listening test hears 4 sentences for a given condition, for a total of 128 votes per condition which were averaged together. The noise suppressed signals evaluated by the human subjects were based on Talker Set 2, and a minor variation of the noise suppressor which is indistinguishable by most human listeners (7-point CCR subjective test shows no preference to within 0.2 MOS points. If there were an effect of this noise suppressor variation, it would be to slightly decrease the Subjective Scores. Thus, switching to the strictly correct variation would have the effect of increasing the Subjective Scores slightly, thus slightly magnifying the differences to the 3Quest predictions.)

3. Correlation Results – ETSI EG 202 396-3 (3QUEST)

In Figure 1, we show the correlation results for ETSI EG 202 396-3 (3Quest) for the 72-point sweep of 6 SNRs and 12 Noise Suppression values, for the S-MOS, N-MOS, and G-MOS scores.

This broad-ranging sweep of SNRs required us to manipulate the noise levels relative to the nominal calibrated level at which the “pub” babble recording was made. ETSI EG 202 396-3 was trained and tested only on noise sources at their nominal calibrated level, so our 72-point sweep represents a test of ETSI EG 202 396-3 outside its validated range. However, we have determined that the SNR=18-dB sweep condition is directly comparable to testing a candy-bar phone like the Google Nexus One in the 3Quest café noise condition, with the speech reduced by 0.7 dB, based on the analysis given in Appendix I. So, it is still reasonable to use this 12-point subset of the 72-point sweep to evaluate the predictive performance of ETSI EG 202 396-3, since it is held constant at its nominal, calibrated background noise level.

Figure 2 shows the corresponding SNR=18dB subset of the data.

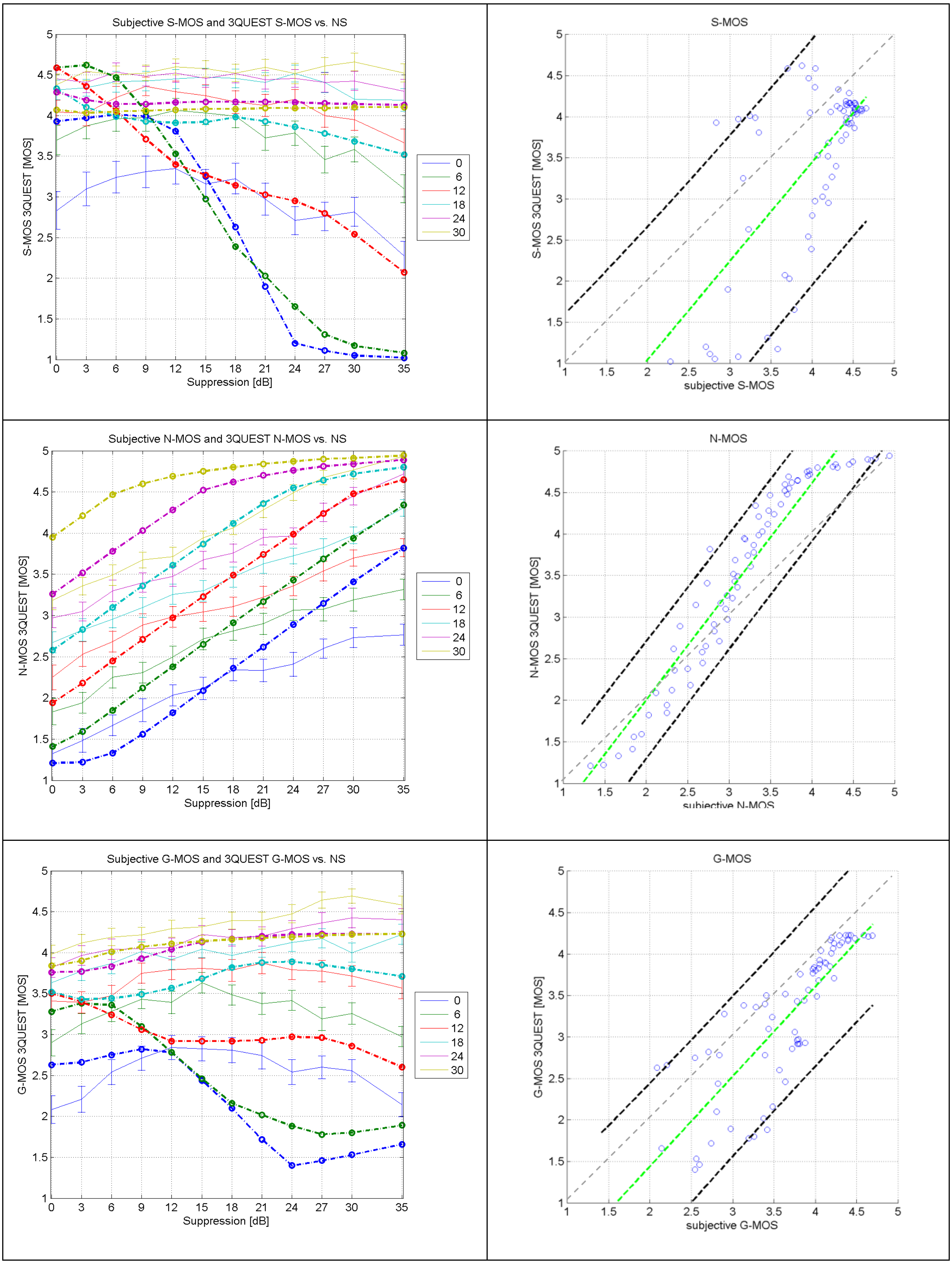


Figure 1: Comparison of 3Quest S-MOS, N-MOS, and G-MOS scores to P.835 scores, for SNRs ranging from 0 to 30dB. P.835 subjective scores are thin lines with error bars. 3Quest scores are thick dashed lines. Best linear regression is a green dashed line. 95% confidence interval is black dashed lines. Perfect correlation reference is gray dashed line. Experiment was done with Talker Set 1, Noise Suppressor 1.

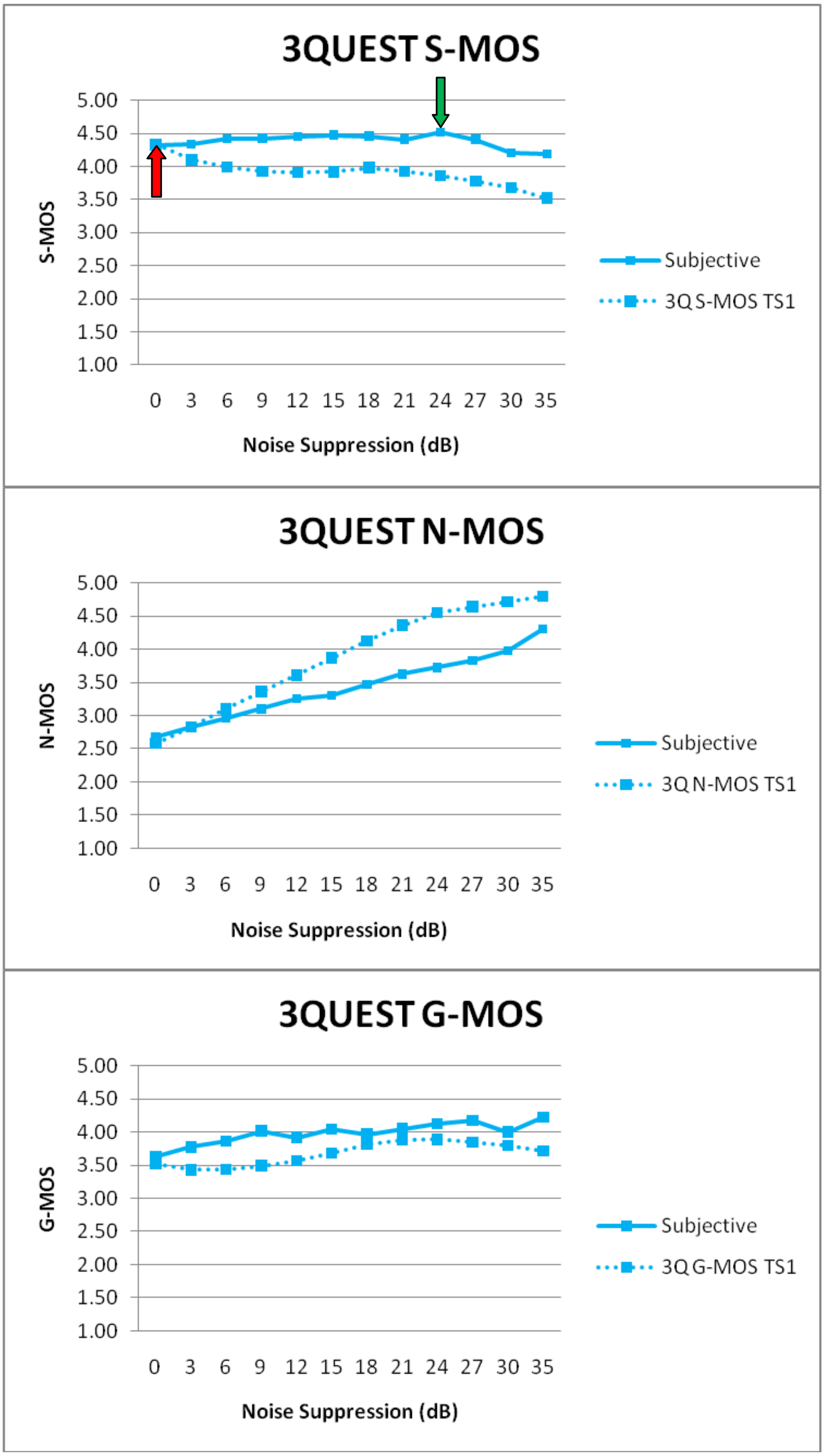
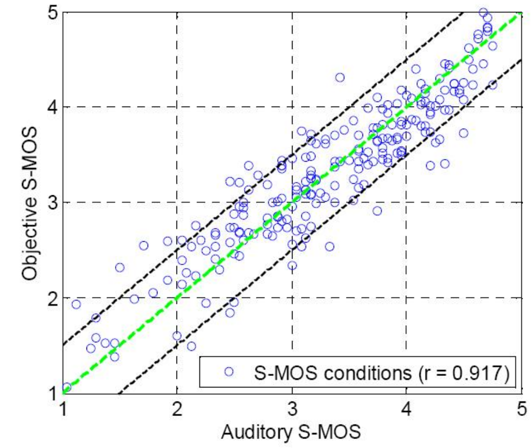
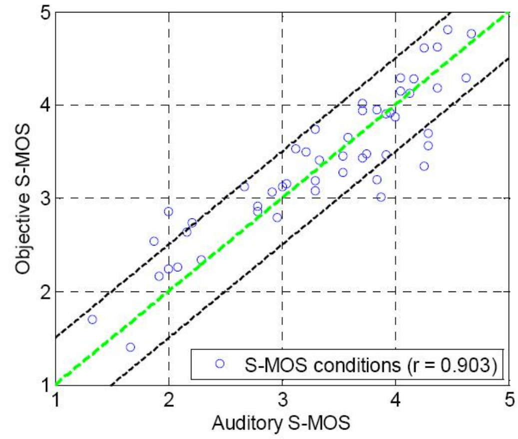


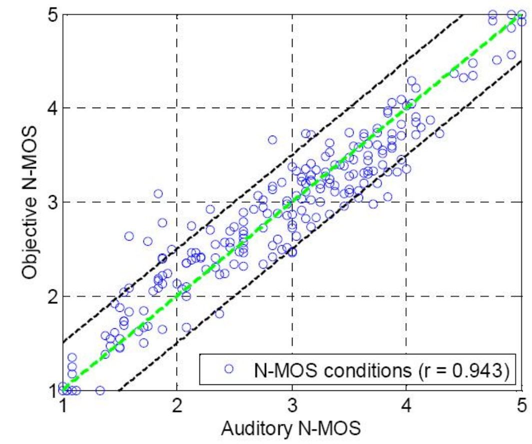
Fig. 2: Comparison of P.835 subjective Scores and EG 202 396-3 3Quest predictions for the SNR = 18dB case. TS1, NS1 designates that this experiment was done with Talker Set 1, as described in Section 2. Subjective and Objective optima for S-MOS are shown with Green and Red arrows, respectively. Note that even though the errors are within 0.7 MOS, there is a 24dB difference in the optimal Noise Suppression values.

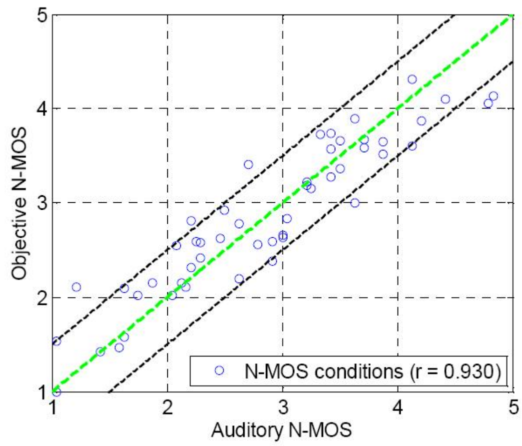
For reference, we also provide the original training and validation data from ETSI 202-396-3 Annex H:

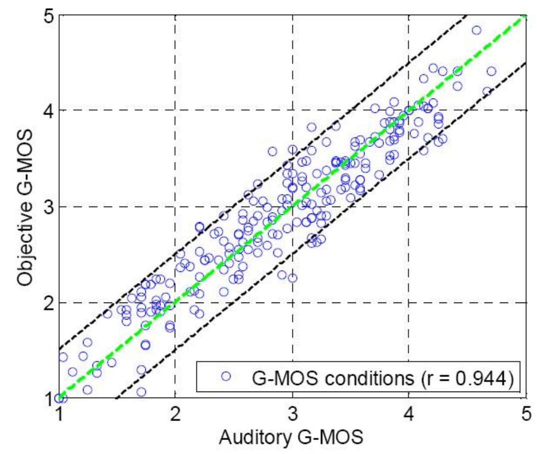
Training Results	Validation Results











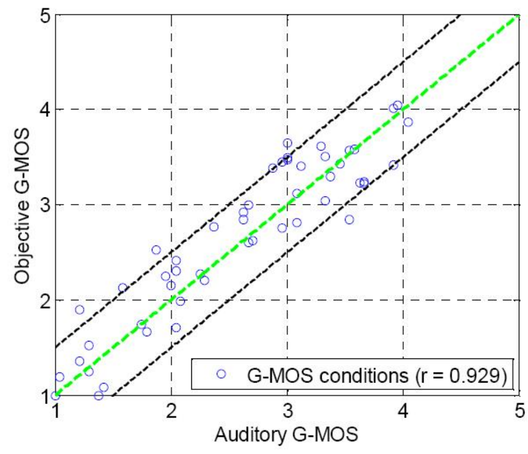


Fig. 3: Original narrowband performance from ETSI 202-396-3 Annex H.

4. Discussion

4.1 Prediction Accuracy over Wide SNR Range (0-30dB)

We make the following observations about the data shown in Figure 1:

1. **N-MOS, SNR = 0-30dB:** In Figure 1, the N-MOS predictions all show the correct general trend, of monotonically increasing N-MOS with Suppression, as expected. There is a tendency to over-predict N-MOS by up to 1.0 MOS points at high SNRs and high suppression values. The errors are within a range of about ±1.0 MOS points, consistent with both the training and validation results of Figure 3.
2. **S-MOS, SNR = 0-30dB:** In Figure 1, the S-MOS predictions have errors with magnitude up to 2.5 MOS points, much higher than the original training and validation data of Figure 3. The general trend is to under-predict S-MOS by 0.5-1.0 MOS points (green line below dashed gray line).
3. **G-MOS, SNR = 0-30dB:** In Figure 1, the G-MOS predictions have errors with magnitude up to 1.7 MOS points, much higher than the original training and validation data of Figure 3. The general trend is to under-predict G-MOS by 0.5-0.7 MOS points (green line below dashed gray line).
4. **Overall:** Generally, only N-MOS appears to be reasonably well predicted over the full sweep range of 0-30dB SNR; The S-MOS and G-MOS predictions both show large errors and a bias toward under-prediction.

ETSI EG 202-396-3 was trained and validated only on the single, fixed calibrated level of the noise files in the ETSI EG 202-396-1 database; and thus, many of the points of our 72-point sweep are outside the trained and validated scope of operation of ETSI EG 202-396-3.

However, P.835 Appendix III Amendment I (10/2007), which pre-dates ETSI EG 202-396-3 (11/2008) by over a year, specifies that SNR values be swept over a 0, 6, and 12dB range for all noise types. And there are commercial wireless carriers who are specifying single SNR levels that do not match the ETSI EG 202 396-1 nominal levels. In practice, modern noise suppression systems must adapt their suppression strengths as a function of SNR, so it necessary to test that this function has been performed correctly at a variety of SNRs. So, while sweeping SNR and non-standard SNRs may be outside the original scope of ETSI EG 202-396-3, in practice a valid predictive tool that can predict P.835, App. III [4] scores must be able to accommodate testing with a sweep of SNR and non-standard SNRs relative to the ETSI EG 202-396-1 database. The limited range of operation of ETSI EG 202-396-3 would appear to place a severe limitation on the practical application of the tool.

In practice, we see that many companies are attempting to use ETSI EG 202-396-3 in SNR sweeps over 0-12dB ranges, since they are collecting P.835, App. III [4] compliant subjective data at significant expense over that range, and want to use that existing data to validate ETSI EG 202-396-3, and are unable to do so. For example, Orange has described their attempts to correlate their P.835 data with ETSI EG 202-396-3 in the 3GPP reports TSG-SA4 #62 S4 (11) 0085 and TSG-SA4 #63 S4-110277.

4.2 Prediction Accuracy at Nominal SNR Level (18dB)

We make the following observations about the data shown in Figure 2:

1. **N-MOS, SNR=18dB:** In Figure 2, the N-MOS prediction shows the correct general trend, of monotonically increasing N-MOS with Suppression, as expected. The errors are within a range of about ±0.8 MOS points, consistent with both the training and validation results of Figure 3.
2. **S-MOS, SNR=18dB:** In Figure 2, the S-MOS predictions are within 0.7 MOS point of the

subjective S-MOS scores over the full range of noise suppression tested. In Figure 2, note that even though the S-MOS predictions are within 0.7 MOS point of the subjective S-MOS scores, the predicted optimal Noise Suppression value is 24 dB different than the subjective optimal Noise Suppression value.

3. **G-MOS, SNR=18dB:** In Figure 2, the G-MOS predictions are within 0.5 MOS point of the subjective S-MOS scores over the full range of noise suppression tested.
4. **Overall:** All three scores are predicted with moderate absolute errors within 0.8 MOS points. In this range, the performance is consistent with the performance reported in ETSI 202-396-3 Annex H and repeated in Figure 3.

The fact that the S-MOS predictions can be within the generally accepted range (0.7 MOS) and still give such a large difference in predicted optimal suppression value (24dB) calls into question whether the generally accepted range is tight enough, given the way the tool will be used in practice. Even though ETSI EG 202-396-3 is not specifically recommended or endorsed to be used to tune or optimize performance of a noise suppressor, the fact that it is being seriously considered as an acceptance test in 26.132 CR (3GPP TSG-SA4 #65 S4-110655 [8]) will necessarily drive handset makers and noise suppression vendors to have to adjust their noise suppression parameters to meet performance levels based on ETSI EG 202-396-3, if the 26.132 CR is approved in its current form. We are very concerned that this may result in potentially drastic mistuning of the devices as illustrated in Figure 2, so as to get good ETSI EG 202-396-3 scores.

4.3 Talker Dependence

The results shown in Figures 1 and 2 were collected with Talker Set 1, as described in Section 2, and represent the best performance we saw for ETSI 202-396-3 on ten different experiments. Figure 4 shows a comparison of the performance difference when running the noise suppressor and ETSI 202-396-3 predictor using Talker Set 1 and Talker Set 2.

Figure 4 shows that N-MOS was completely insensitive to the change between Talker Set 1 and Talker Set 2, as expected, but S-MOS and G-MOS were strongly affected by the difference between the talker sets, which was not expected. The worst-case difference introduced by the Talker Set difference was 1.4 MOS points for S-MOS, and 1.0 MOS points for G-MOS, a very large effect.

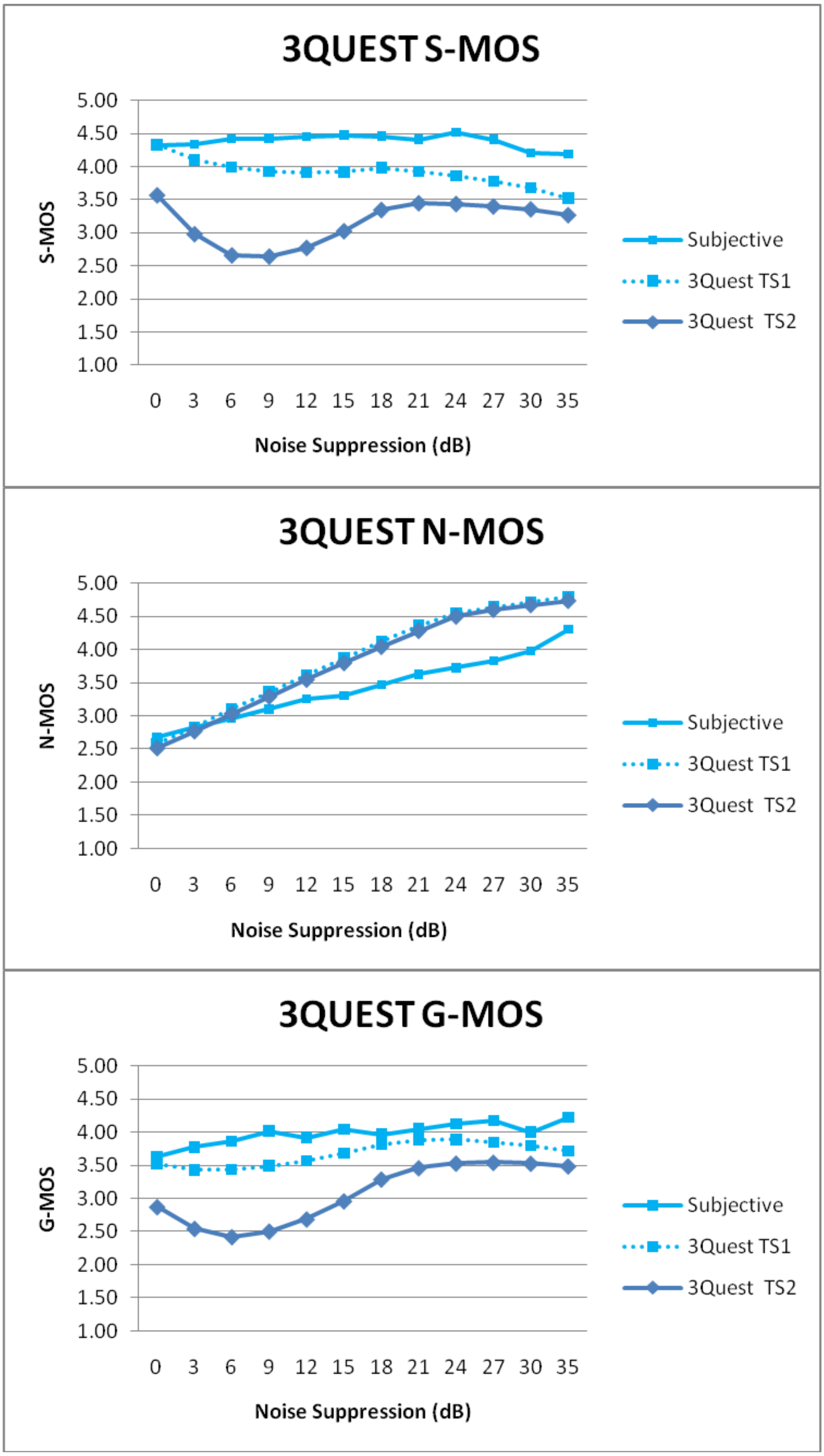


Fig. 4: Comparison of 3Quest performance for Talker Set 1 (dotted light blue line) and Talker Set 2 (solid dark blue line).

4.4 Other Factors

There are other differences between the current study and the original study, as listed below:

	Original Study	Current Study
Panel Nationality	German	American
Number of microphones For Training	All 1-mic	All 1-mic
Number of Microphones For Validation	All 1-mic (matches training data)	All 2-mic (does not match training data)

However, we have not needed to invoke the "panel nationality" or "number of microphones" factors to explain the observations made in the current study.

4.5 Extensions

The experiments above only considered the babble distracter and used only a simple noise suppressor – suppression only, no canceller component, constant suppression (no time-varying suppression behavior), no AGC. All of those features are common in modern 2011-vintage noise suppressors.

An extension to additional distracters and noise reduction systems that include both cancellation and suppression was undertaken. A series of experiments was conducted for seven distracters, including the six listed in P.835 App III [4], pink, car, traffic, babble, music, voice, plus an additional train station distracter. Three distracters, car, traffic, and train station, were taken from the ETSI EG 202-396-1 database (Fullsize_Car1_130kmh_binaural, Outside_Traffic_Road_binaural, and Train_Station_binaural). The music distracter consisted of a 30-sec sample of 'rock'-style music containing electric guitar and drums, but no vocal. The voice distracter consisted of alternating male and female talkers uttering short sentences. Pink noise was produced using uncorrelated sources from each of the four loudspeakers.

For each distracter, the SNRs tested were 0, 6, 12, and 24dB. Two noise reduction systems consisting of two-microphone hybrid canceller/suppressor architectures, with close (2-cm) and far (8-cm) Mic Distances (MD) controlling the degree of cancellation, were each tested over a range of Noise Suppression (NS) values of 0, 6, 12, 18, 24, and 30dB. The suppressor component was the same as described in Section 2.

For each distracter, the total number of conditions was 4 SNR x 6 NS x 2 MD = 48 in total. These were each presented to separate panels of 32 naïve listeners in a P.835 methodology, as described above. In all, seven listening panels were used for the extensions described in this section.

The reference system used for these extensions departs from the traditional usage of Modulated Noise Reference Unit (MNRU) to degrade the subjective S-MOS dimension. A controlled degradation of the speech based on a Wiener filter model was used as a replacement for MNRU. This approach and validation results are described in a contribution to the ITU-T SG-12 Q7 Rapporteur's meeting of June 2011 [7].

Results are shown below, as functions of NS, similarly to the left panels of Figure 1, where thin lines with error bars give subjective results, and thick lines with open symbols give 3Quest prediction. Predictions are for Talker Set 2, which was shown in Section 4.3 to impart lesser talker dependence to the 3Quest predictions. Results for S-MOS are presented in Figure 5, results for N-MOS in Figure 6, and results for G-MOS in Figure 7.

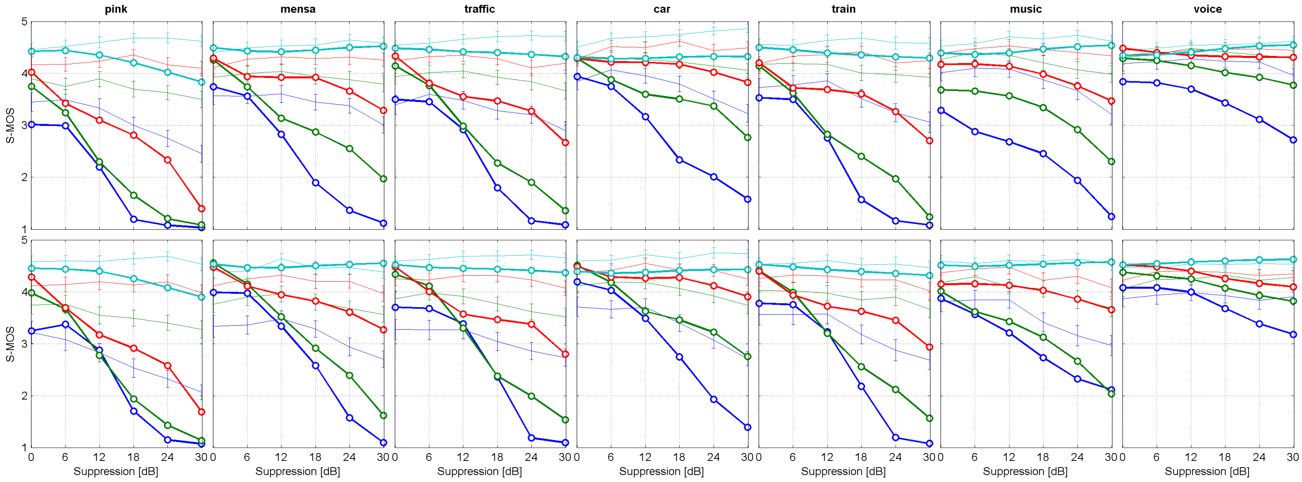


Fig. 5: S-MOS results for hybrid noise reduction systems. Upper row is for close-mic (2cm) spacing; lower row is for far-mic (8cm) spacing. Each column is for a different distracter. Thin lines are subjective data, thick lines are predictions. Blue color for SNR=0dB, green color for SNR=6dB, red color for SNR=12dB, and cyan color for SNR=24dB.

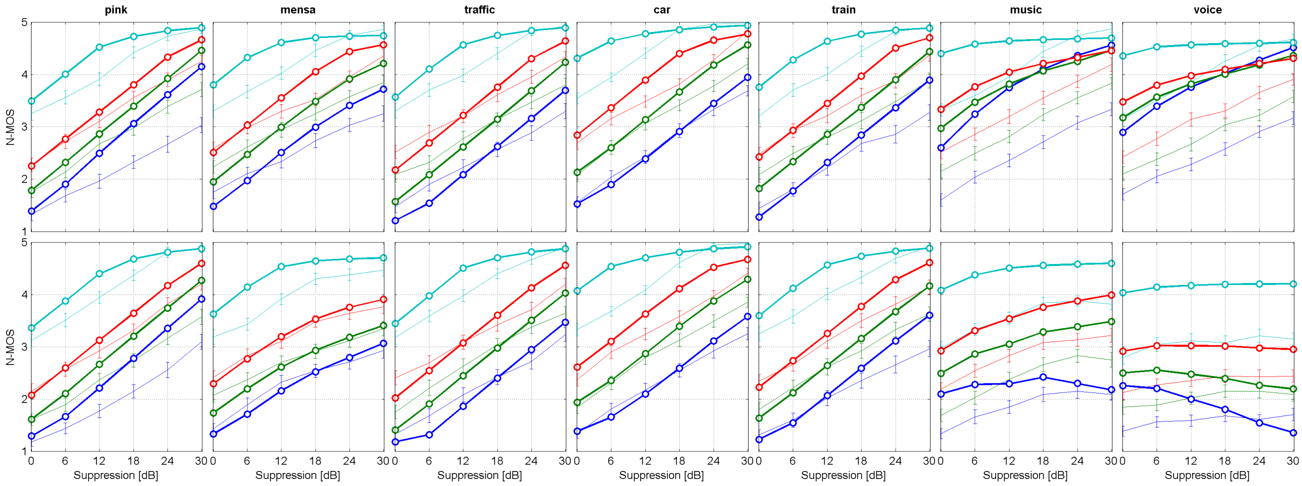


Fig. 6: N-MOS results for hybrid noise reduction systems. Arrangement and coding by color and line as for Figure 5.

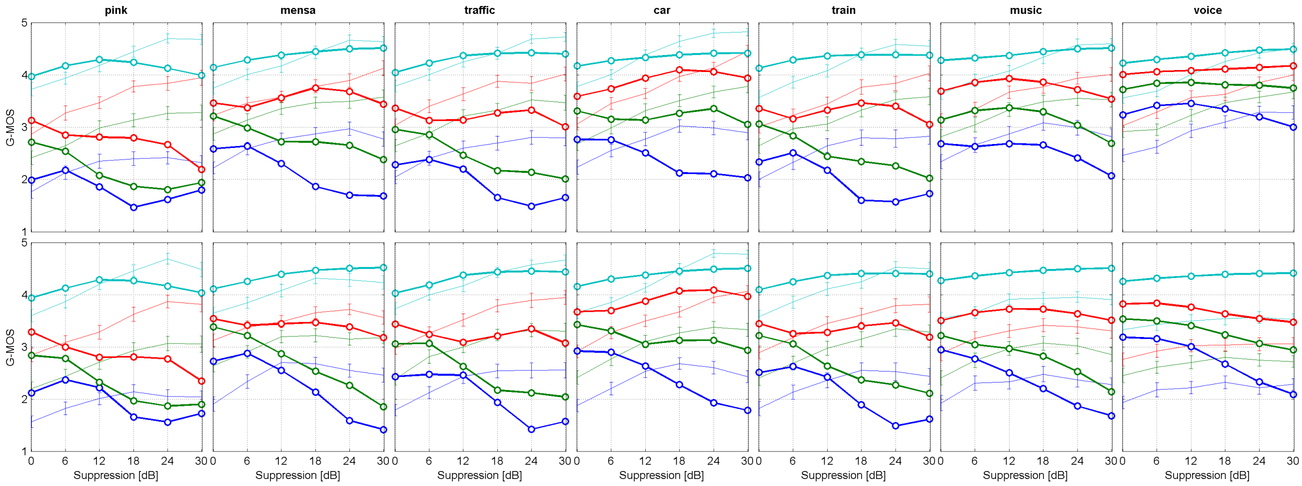


Fig. 7: G-MOS results for hybrid noise reduction systems. Arrangement and coding by color and line as for Figure 5.

In general, the trends observed in Section 3 are also observed in these results. Note that the N-MOS predictions show relatively higher error for the music and voice distracters. Corresponding scatter plots are shown in Figures 8, 9, and 10.

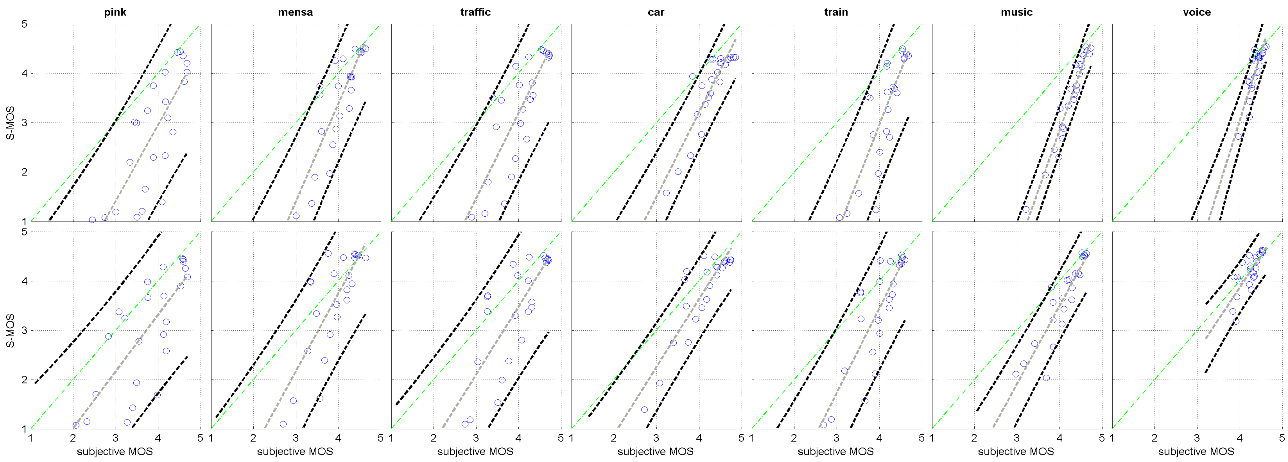


Fig. 8: S-MOS scatter plots for hybrid noise reduction systems. Upper row is for close-mic (2cm) spacing; lower row is for far-mic (8cm) spacing. Each column is for a different distracter. Subjective scores are plotted on the abscissa, corresponding predictions plotted on the ordinate. The green dashed line represents one-to-one mapping; the grey dashed line represents best linear fit; dark dashed lines represent 95% confidence intervals on the best linear fit.

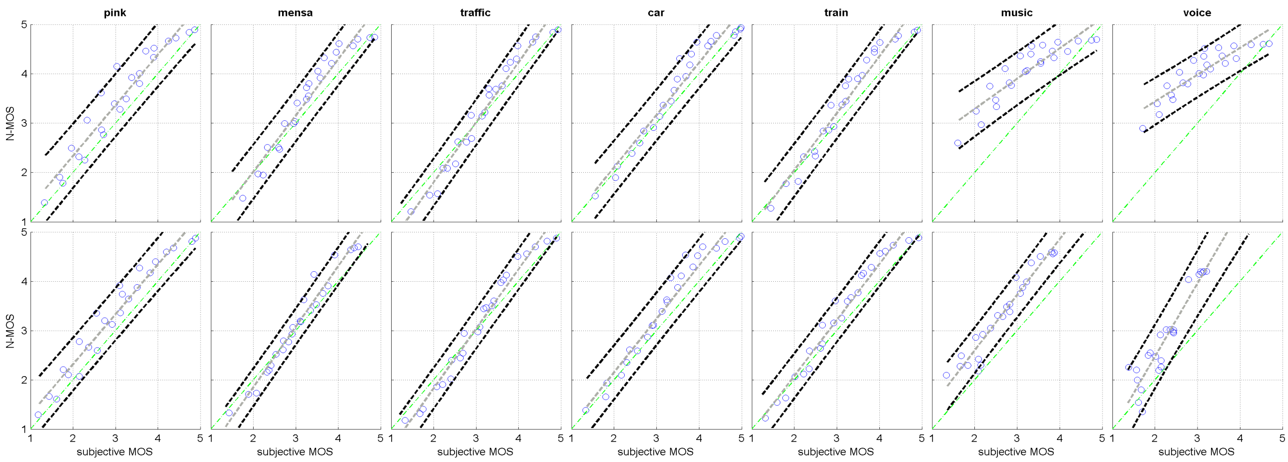


Fig. 9: N-MOS scatter plots for hybrid noise reduction systems. Arrangement and coding of lines is as for Figure 8.

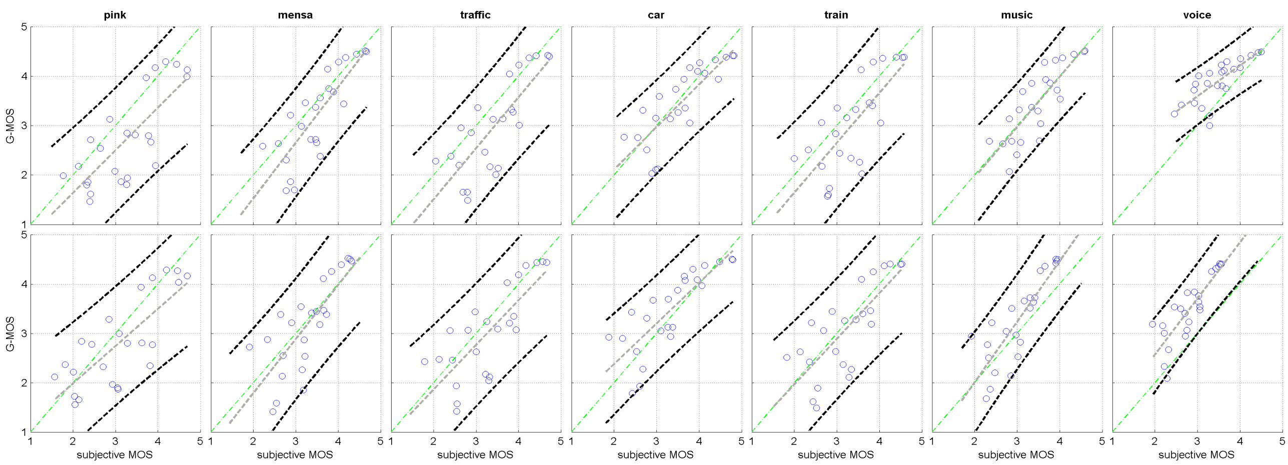


Fig. 10: G-MOS scatter plots for hybrid noise reduction systems. Arrangement and coding of lines is as for Figure 8.

5. Summary and Conclusions

Based on the data presented in this study, we offer the following summary of our findings:

1. Generally, the performance of ETSI EG 202-396-3 is within the originally published ranges for N-MOS across the full range of SNRs and Noise Suppression strengths. N-MOS predictions also showed no Talker dependence.
2. The valid range of operation of ETSI EG 202-396-3 for S-MOS and G-MOS is limited to the fixed SNRs corresponding to the fixed, calibrated noise levels of ETSI EG 202-396-1 for each distracter; outside this range, for SNRs = 0, 6, and 12dB, prediction errors have been observed to be very large (2.5 MOS for S-MOS, 1.7 MOS for G-MOS). Since the currently in-force version of P.835, App. III [4] specifies SNRs of 0, 6, and 12dB, this limitation of ETSI EG 202-396-3 would appear to preclude it from predicting the currently in-force version of P.835, App. III [4]. Even when used with the approved SNRs, the predicted optimum suppression value may be incorrect by as much as 24dB, and thus great care should be taken to ensure that ETSI 202-396-3 should not be endorsed explicitly or implicitly for use as an optimization tool, which would be in inescapable consequence if it is approved as an acceptance test in 3GPP TSG-SA4 #65 S4-110655.
3. ETSI EG 202-396-3 exhibits a strong talker and sentence dependence for S-MOS and G-MOS, which indicates that it is limited to only using the P.501 samples on which it was trained and validated; with other speech samples, S-MOS and G-MOS prediction errors have been observed to be very large (1.4 MOS for S-MOS, 1.1 MOS for G-MOS).

References

[1] ITU-T Recommendation P.835: Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm (11/2003)

[2] ETSI EG 202 396-3: Speech Quality performance in the presence of background noise Part 3: Background noise transmission - Objective test methods (11/2008)

[3] ETSI EG 202 396-1: Speech Quality performance in the presence of background noise Part 1: Background noise simulation technique and background noise database (03/2009)

[4] ITU-T Recommendation P.835: Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm, Amendment 1: New Appendix III - Additional provisions for nonstationary noise suppressors", ITU-T Recommendation P.835 Amendment 1 (10/2007)

[5] 3GPP TSG-SA4 #62 S4 (11) 0085, Correlation of Objective Speech Quality Measures according to ETSI EG 202 396-3 to P.835 Subjective Test Results

[6] 3GPP TSG-SA4 #63 S4-110277, Correlation of Objective Speech Quality Measures according to ETSI EG 202 396-3 to P.835 Subjective Test Results: Additional Results.

[7] ITU-T SG-12 Q7 Rapporteur’s Meeting document AH-11-029, Better Reference System for the P.835 SIG Rating Scale. Joint contribution of Dynastat and Audience.

[8] 3GPP TSG-SA4 #65 S4-110655, Draft Change Request to TS 26.132 on Extension of Acoustic Tests (Release 11).

Appendix I: Comparison of Sweep Conditions to 3QUEST Training / Validation Conditions

Information on the noise files and speech levels was provided by Head Acoustics to Audience. Information on the path loss (MRP-to-Primary Mic) for four phones was provided by the Audience QE team [page 2]. Based on this information, for the Nexus 1 phone, the Audience condition of 'babble, 18 dB SNR at primary mic' condition is within 0.7 dB of the café noise condition used in 3Quest training (café SNR at MRP = 27.8 dB, babble-18 dB SNR referred to MRP = 27.1 dB).

We find that the 18-dB sweep condition is directly comparable to testing the Nexus 1 phone in the 3Quest café noise condition, with the speech reduced by 0.7 dB.

Note that the Audience babble condition uses the **pup** noise file from ETSI EG 202 396-1, while the analogous condition for 3Quest uses the **café** noise file. The principle difference is that the **café** file includes impulses (dishware) and also has short intervals where a single voice (laughter) dominates [see time plots pages 3 & 4]. Comparing power spectra [Figure 10] shows that **café** has somewhat more energy below 200Hz (HVAC?) and above 3000Hz (dishes or other?) than does **pup**, which seems to be more uniformly multi-talker babble.

Head Acoustics 3QUEST Narrowband development & validation conditions								
	EG 202 396-1 Filename	Duration		Documented	Measured		SNR @	**SNR @**
Name		sec		dB(A)	dB(A)	dBSPL	MRP (A)	**MRP (A)**
cal	Sine0dBPa_1sec	1	L	94	94.0	94.1		avg
			R	94	94.0	94.1		
café noise*	Mensa_binaural	22	L	63.4	63.4	83.1	27.1	**27.8**
			R	61.9	61.9	82.7	28.6	
car noise	Fullsize_Car1_130kmh	30	L	69.1	69.2	98.2	21.3	**21.9**
			R	68.1	68.1	99.4	22.4	
train station	Train_Station_binaual	30	L	68.2	68.2	84.4	22.3	**21.5**
			R	69.8	69.8	84.8	20.7	
street noise	Outside_Traffic_Road_binaural	30	L	74.9	74.9	79.7	15.6	**16.1**
			R	73.9	73.9	79.7	16.6	
	*Measured level differs from published in EG 202 396-1. File date is 11 May 2009.							
	files from http://docbox.etsi.org/stq/Open/EG%20202%20396-1%20Background%20noise%20database/Binaural							
							MRP**	
					dB(A)	dB SPL	dB(A)	dB SPL
Speech	Speech e_3quest1	39			92.7	94.5	90.5	**92.3**
	**Per F. Kettler: Speech @ MRP = -4.7dBPa + 3dB, for Lombard effect. Net is 92.3dBSPL.							
	***Hand-Held Speakerphone, d_{HF} per TS 26.132 v9.1.0, From Note contained in Clause 5.1.3.3.							

Audience CT (handset) test conditions, from ITU-T P.835, Appendix III								
	from \\fs\Audio\Reference\Distractors\P835							
	EG 202 396-1 Filename	Duration		Level	Measured			SNR @
Name		sec		dB(A)	dB(A)	dBSPL		pri mic (A)
babble	*pub noise binaural	28	L	77.8	70.8	72.8		0, 6, 12
			R	78.9	78.9	81.2		
car	Midsize_Car1_130kmh	30	L	67.0	67.0	97.0		0, 6, 12
			R	65.9	65.9	97.7		
street	Outside_Traffic_Road_binaural	30	L	74.9	74.9	79.7		0, 6, 12
			R	73.9	73.9	79.7		
pink	not in EG 202 396-1	30	L	na	91.8	96.6		0, 6, 12
			R	na	91.8	96.6		
music	not in EG 202 396-1	28	L	na	91.3	93.4		0, 6, 12
			R	na	91.3	93.4		
voice	alternating male/female	49	L	na	81.7	87.1		0, 6, 12
			R	na	82.4	87.9		
	*older pub_noise_binaural, Not V2. Level taken from V1.1.2 (2006-1) of EG 202 396-1							

			Difference between HA and Audience SNRs at MRP:															
			Phone A	4.3			Phone B	7.9			Phone C	9.1			Phone D	8.0		
	SNR at pri mic		0	6	12	18	0	6	12	18	0	6	12	18	0	6	12	18
	SNR at MRP		4.3	10.3	16.3	22.3	7.9	13.9	19.9	25.9	9.1	15.1	21.1	27.1	8.0	14.0	20.0	26.0
HA 3QUEST	Café	27.8	23.5	17.5	11.5	5.5	19.9	13.9	7.9	1.9	18.7	12.7	6.7	0.7	19.8	13.8	7.8	1.8
	Car	21.9	17.6	11.6	5.6	-0.4	14.0	8.0	2.0	-4.0	12.8	6.8	0.8	-5.2	13.9	7.9	1.9	-4.1
	Train	21.5	17.2	11.2	5.2	-0.8	13.6	7.6	1.6	-4.4	12.4	6.4	0.4	-5.6	13.5	7.5	1.5	-4.5
	Street	16.1	11.8	5.8	-0.2	-6.2	8.2	2.2	-3.8	-9.8	7.0	1.0	-5.0	-11.0	8.1	2.1	-3.9	-9.9

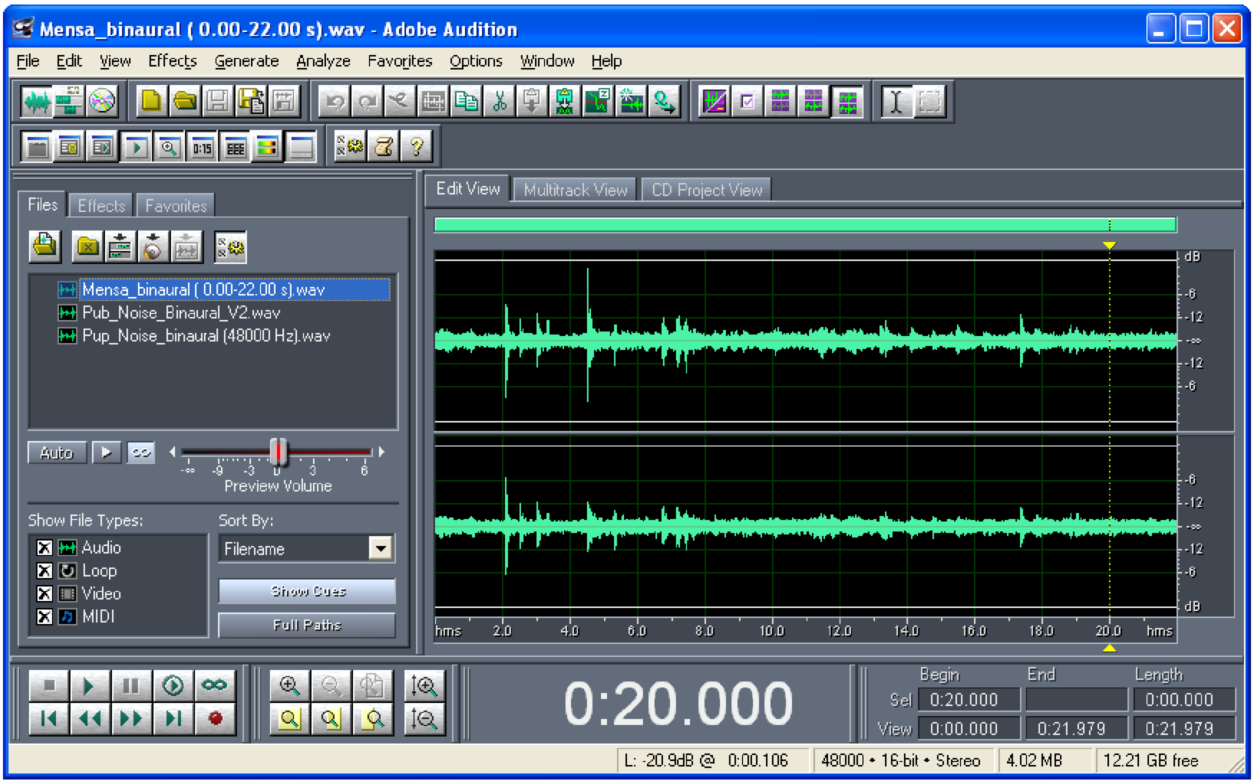


Figure 7: Time plot for Mensa_binaural file.

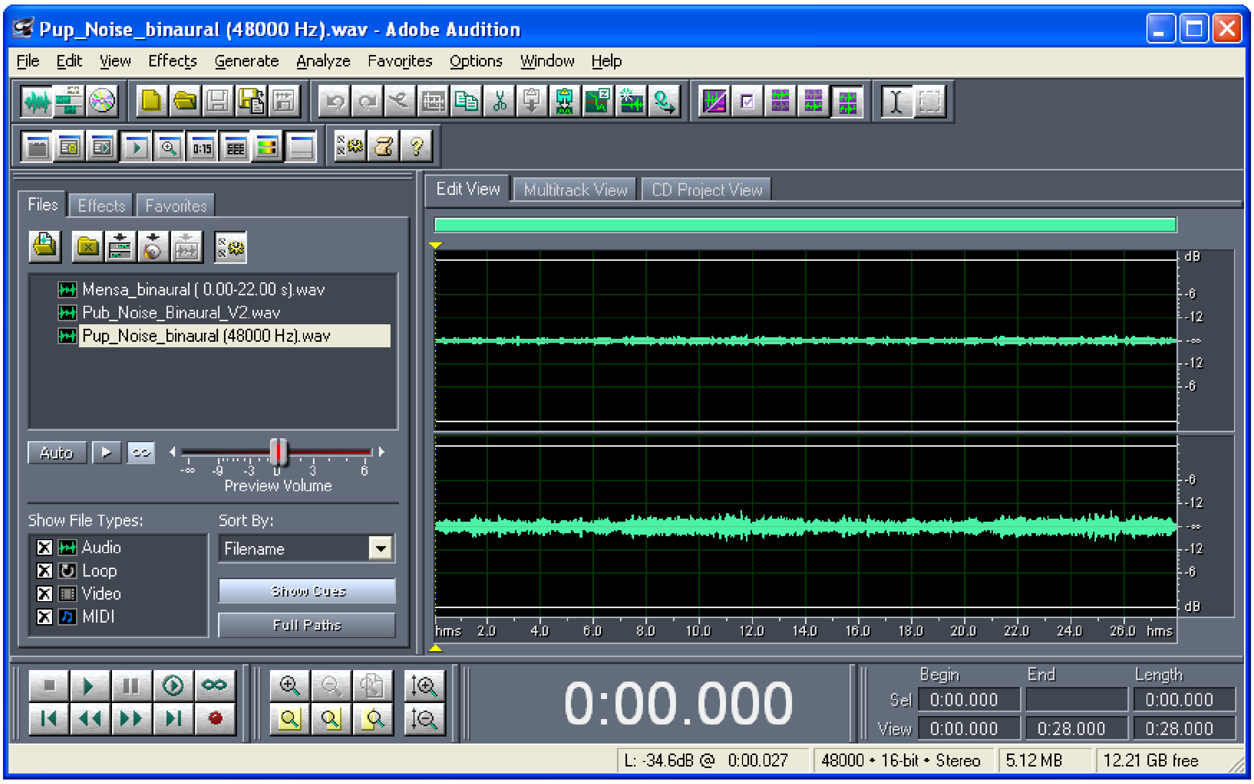


Figure 8: time plot for Pup_Noise_binaural file.

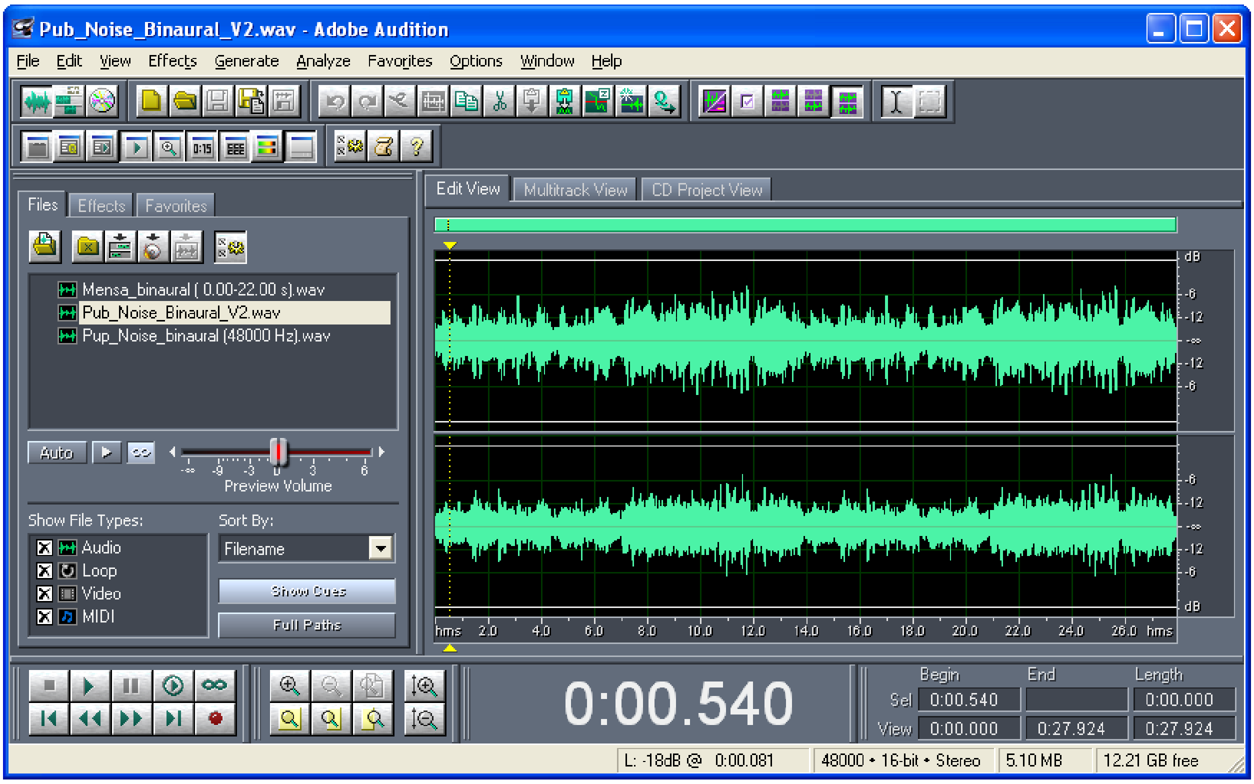


Figure 9: Time plot for Pub_Noise_Binaural_V2 file.

Note: 'Pu**p**_Noise_binaural' is an earlier version that 'Pu**b**_Noise_Binaural_V2' replaces.

The differences are: levels are more closely matched and overall higher in Pub, but appear to be scaled versions of Pup; the loop point is cross-faded in Pub, whereas there is a short silence in Pup.

While Audience uses the older Pup file, the levels in each channel are set so as to be similar as in Pub_v2 [not shown in the table, which lists the levels in-file, and the target SNRs]. Also, loop points are smoothed and extended for the longer speech files.

In practical terms, this could be considered as nearly identical to using the current Pub_v2.

Figure 10: Power Spectra of noise files: Green=Mensa_binaural; Red=pup_binaural; Blue=pub_binaural_V2.