IEEE Transactions on Speech and Audio Processing, July 1999

Objective Estimation of Perceived Speech Quality, Part II: Evaluation of the Measuring Normalizing Block Technique

Stephen Voran

Institute for Telecommunication Sciences
U.S. Department of Commerce, NTIA/ITS.N3
325 Broadway
Boulder, Colorado 80303
sv@bldrdoc.gov
303-497-3839 Voice
303-497-5323 Fax

EDICS Number: SA 1.4.6

Abstract

Part I of this paper describes a new approach to the objective estimation of perceived speech quality. This new approach uses a simple but effective perceptual transformation and a distance measure that consists of a hierarchy of measuring normalizing blocks. Each measuring normalizing block integrates two perceptually transformed signals over some time or frequency interval to determine the average difference across that interval. This difference is then normalized out of one signal, and is further processed to generate one or more measurements. In Part II the resulting estimates of perceived speech quality are correlated with the results of nine subjective listening tests. Together, these tests include 219 4-kHz bandwidth speech codecs, transmission systems, and reference conditions, with bit rates ranging from 2.4 to 64 kb/s. When compared with six other estimators, significant improvements are seen in many cases, particularly at lower bit rates, and when bit errors or frame erasures are present. These hierarchical structures of measuring normalizing blocks, or other structures of measuring normalizing blocks may also address open issues in perceived audio quality estimation, layered speech or audio coding, automatic speech or speaker recognition, audio signal enhancement, and other areas.

I. Introduction

Part I of this paper describes a new approach to the objective estimation of perceived speech quality. This approach uses a simple but effective perceptual transformation, and a hierarchy of measuring normalizing blocks (MNBs) to compare perceptually transformed speech signals. In this second part of the paper, we evaluate these new objective estimators of perceived speech quality by comparison with the results of nine subjective tests. Together, these tests include 219 4-kHz bandwidth speech codecs, transmission systems, and reference conditions, with bit rates ranging from 2.4 to 64 kb/s. When compared with six other estimators, the MNB-based estimators show significant improvements in many cases, particularly at lower bit rates, and when bit errors or frame erasures are present. Benchmark objective estimates of perceived speech quality for standardized codecs are also provided.

II. Correlation with Subjective Test Results

The MNB structures defined in Part I of this paper produce auditory distance (AD) values by forming a linear combination of MNB measurements stored in the vector **m**:

$$\mathrm{AD} = \mathbf{w}^{\mathrm{T}} \cdot \mathbf{m}. \tag{1}$$

These values are then passed through a logistic function to create L(AD). The logistic function is

$$\mathrm{L}(z) = \frac{1}{1+\mathrm{e}^{a \cdot z + b}} \ . \tag{2}$$

L(AD) values range from 0 to 1 and are positively correlated with perceived speech quality.

To judge the usefulness of the L(AD) values as estimators of relative perceived speech quality, we have compared L(AD) and six other established objective estimators of speech quality with the results of formal subjective tests. Nine absolute category rating (ACR) tests that use the 5-point mean opinion score (MOS) scale tests were available to us, and they are summarized in Table I. While the objective estimator structure more closely parallels degradation category rating (DCR) subjective tests, only ACR subjective tests were available for this study. Together, these nine tests include 219 4-kHz bandwidth speech codecs, transmission systems, and reference conditions, with bit rates ranging from 2.4 to 64 kb/s, and some analog conditions as well. Both flat and intermediate reference system (IRS) [1] filtered speech material was included. (IRS filtering simulates the sending response of a typical telephone handset.) A total of 22 hours of speech from at least 52 different speakers, both male and female, in three different languages was used. This collection of speech files and scores has allowed us to complete one of the most comprehensive tests of objective estimators of perceived relative speech quality.

Note that the nine formal subjective tests were provided to us by a variety of well-established subjective testing laboratories. Even so, subjective scores from different tests are not necessarily comparable unless the tests share enough common reference conditions to allow for the calculation of calibration factors. This fact is of little consequence here however, since the correlations described below are invariant to both the scaling and the shifting of subjective scores.

The six established estimators are SNR [2], SNRseg [2], perceptually weighted SNRseg (PWSNRseg) [3], cepstral distance (CD) [2], Bark spectral distortion (BSD) [4], and noise disturbance (ND) which is the output of the perceptual speech quality measure (PSQM) algorithm that is defined in the main body of ITU-T Recommendation P.861 [5],[6]. To create a uniform comparison, each of these estimators was passed through the logistic function in (2). For each estimator, the constants *a* and *b* were selected to maximize the coefficient of correlation between the logistic function output and MOS across the nine subjective tests. The maximizing values of *a* and *b* are shown in Table II. The resulting coefficients of correlation are shown in Table III. Pearson correlation is used throughout this paper.

The correlation values in Table III were calculated after averaging all available subjective scores for each condition of a given test to a single score for that condition. Similarly, for each condition of a given test, all available objective estimates were averaged to generate a single objective estimate for that

condition. Thus, we refer to these correlation values as "per-condition" correlations. A more advanced analysis technique, described in [7] recognizes the importance of the distributions of the objective estimates and the subjective scores for each condition, how they influence confidence intervals, and in turn, the final conclusions that one draws from objective and subjective tests.

Table III demonstrates again the limitations of SNR, SNRseg, and PWSNRseg as estimators of perceived speech quality. CD and BSD tend to show higher correlations for tests 5, 6, and 7, which contain only conditions that tend to preserve waveforms. The PSQM result, L(ND) appears to be the most reliable of these six existing objective estimators, across these nine tests. Since tests 1-4, 8, and 9 contain conditions that are outside of the defined scope of the PSQM algorithm, we conclude that the PSQM algorithm can sometimes make useful estimates outside of its scope. Because the PSQM result, L(ND), appears to be the most reliable of these six objective estimators, we use it as the reference against which to compare L(AD).

Table IV shows per-condition correlation values for L(AD) as calculated by the two MNB structures. Since L(ND) is used as a reference, that column from Table III is repeated as column 2 of Table IV to allow for easy comparisons. Two versions of the estimators were evaluated. These versions differ only in the values of the weights used in (1) and constants a and b used in (2).

The first version of each estimator was created by optimizing variables in (1) and (2) to maximize correlation between L(AD) and MOS across tests 1 and 2 only. The parameter a in (2) was absorbed into the weights in (1), resulting in 13 or 12 free variables. These variables were used to fit 1226 data points, so the fitting problem was over-determined by an approximate factor of 100. The resulting correlation values are shown in Table IV, columns 3 and 4. These columns show that this limited optimization results in an objective speech quality estimator that generalizes well to the other seven tests. This result is important because it indicates that these estimators do model perception and judgment, rather than inadvertently modeling some specific properties of the conditions in tests 1 and 2.

To create the most effective estimator, one must use all available data. Thus, we created a second version of each estimator by optimizing variables in (1) and (2) to maximize correlation across all nine tests. This involved fitting 11,812 data points, so the fitting problem was over-determined by a factor greater than 900. The resulting correlations are shown in columns 5 and 6 of Table IV. When all nine tests are considered together, MNB structure 2 appears to be slightly more useful than MNB structure 1. Both structures show dramatic improvements over the PSQM result, L(ND) on tests 3, 8 and 9, which contain the lower rate speech codecs, bit error, and frame erasure conditions.

Tables III and IV provide a complete set of correlation results. Those tables can be used to compare the performance of 10 objective estimators across 9 different subjective tests. In addition, we selected four cases that display a wide range of correlation values and generated subjective-objective scatter plots. These plots allow for visual interpretations of per-condition correlation values. Each plot shows an objective estimator vs MOS, using a single point per condition. Figure 1 shows L(BSD) for test 3 where the per-condition correlation, ρ, is 0.368. Figure 2 shows L(ND) for test 3 where ρ=0.793. Figure 3 gives L(AD) using the fully optimized MNB structure 2, also on test 3, with ρ=0.959. Finally, Figure 4 shows L(AD) using the fully optimized MNB structure 1, on test 5, where ρ=0.986.

III. Observations and Discussion

The optimized values of the variables in (1) and (2) are given in Table A-I, found in Part I of this paper. Because the measurements have different variances, the weights do not indicate the relative importance of the measurements. Note that one weight in Table A-I is zero, indicating that the first measurement in MNB structure 2 does not presently provide useful information for this application. We retain this measurement for completeness, and for its potential future utility in this or other applications. In both structures, the first four weights are applied to FMNB measurements taken at the edges of the speech band. For MNB structure 1, the positive value of $\mathbf{w}(1)$ and the negative value of $\mathbf{w}(2)$, indicate that to

maximize estimated speech quality, energy below 250 Hz (outside the telephony speech passband) should be minimized, but only if energy above 250 Hz can be retained. Similarly, the negative value of **w**(3) and the positive value of **w**(4) indicate that energy above 3250 Hz should be minimized, but not at the expense of energy below 3250 Hz. These data-driven mathematical results agree with our intuitions about in-band speech power and out-of-band noise. As part of a sensitivity analysis, we determined that when the weights in **w** are perturbed from their optimal values by 10%, resulting coefficients of correlation are reduced by about 1%. In addition, 1% and 0.1% perturbations in the weights result in 0.1% and 0.01% reductions in correlation, respectively.

Table IV shows that correlations between fully optimized MNB estimators and subjective scores range from .910 to .986. Given the breadth of conditions covered by the nine tests, these are very encouraging results. In particular, the improved ability to estimate perceived speech quality for lower rate speech codecs, some of which are operating with bit errors or frame erasures, represents an important advance. Based on this improvement, ITU-T Recommendation P.861 has been updated by the inclusion of an MNB algorithm in Appendix II of the Recommendation [8]. The algorithm that appears there is an earlier version of MNB structure 2 described in this paper.

For unoptimized software implementations, we found that either of the MNB-based estimators requires approximately 920,000 floating-point operations to process 1 second (8000 samples) of speech. Because the bulk of these operations is devoted to the FFT, both MNB algorithms can be run at the same time using only 940,000 floating-point operations. Similarly, an unoptimized implementation of the PSQM algorithm required about 1.21 million floating-point operations to process 1 second (8000 samples) of speech.

We have also implemented the MNB estimators with the frame overlap reduced from 50% to zero. This reduces the number of computations in the unoptimized implementation by a factor of two but has surprisingly little impact on estimator performance for the conditions described in Table I. When the frame overlap is reduced to zero and the parameters given in Table A-I are optimized, the resulting coefficients of correlation shown in Table IV all change by less than 0.5% from their original values. In spite of this result, we do not recommend implementations with zero frame overlap because the estimator could be extremely vulnerable to certain periodic, frame-synchronous noises and distortions. In addition, 50% overlap of Hamming windows places equal weight on each speech sample but zero overlap does not.

We tested the MNB estimators to determine how sensitive they are to errors in delay estimation. Two groups of 96 speech files were selected for this test. One group had PSDs sufficiently preserved such that fine delay estimates (as described in Part I of this paper) were almost always possible. The second group contained more severely distorted speech files for which fine estimates were rarely possible. We refer to these two groups as the fine group and the coarse group, respectively. Once the actual delays of the files were determined, the files were shifted to create temporal misalignments of 1, 2, 4, 8, 16, and 32 ms in both directions. The original, zero misalignment, L(AD) values for the 192 files used ranged from 0.16 (rather low quality) to 0.97 (very high quality). As expected, the temporal misalignments caused L(AD) to drop. We calculated the percentage drop in L(AD) for each file at each value of temporal misalignment. The percentage drop in L(AD) varied significantly between speech files. When averaged over all 96 files in a group, the results were very similar for the two MNB estimators, and for both time directions. Averaging over the two estimators and the two time directions yields the results in Table V. As expected, L(AD) drops more slowly for the files in the coarse group since an exact delay value does not even exist for those files.

IV. Benchmark Values

Tables VI-IX provide benchmark values of AD and L(AD) for both MNB algorithms. Results are given for 16 standardized speech codecs and for 14 modulated noise reference unit (MNRU) [9] conditions. Within each table, AD and L(AD) results generally agree with known results on the perceived quality of

these codecs and MNRU conditions. These results provide context for AD or L(AD) measurements made on other codecs or conditions.

Each condition in Tables VI-IX was evaluated using a total of 64 English-language sentence pairs. These 64 sentence pairs come from 4 female and 4 male talkers, each providing 8 different sentence pairs. Together, the 64 sentence pairs last about 400 seconds. Two sets of values were computed. Wideband speech recordings were band limited to 200-3400 Hz using a flat bandpass filter and then passed through the 30 conditions listed in the tables. Values for this "flat speech" experiment are given in Tables VI and VII. In addition, wideband speech was filtered according to the IRS sending sensitivity characteristic [1] and then passed through the 30 conditions. Values for this "IRS speech" experiment are given in Tables VIII and IX. The tables provide a mean value taken across all 64 sentence pairs, as well as the half-width of the 95% confidence interval about that mean.

The MNRU is the most common reference condition for subjective and objective speech quality assessments. A common anchoring technique uses MNRU conditions with Q (SNR) values at 5- or 6-dB increments. We have provided benchmark values for MNRU conditions with Q values between 0 and 40 dB, in 5- and 6-dB increments. In addition, (3) through (6) give quadratic fits between Q and AD for the 4 cases that correspond to Tables VI - IX.

$$AD \approx 0.0003 \cdot Q^2 - 0.1862 \cdot Q + 8.1859, \quad 0 \le Q \le 40, \quad \text{MNB} - 1, \text{Flat Speech} \tag{3}$$

$$AD \approx 0.0020 \cdot Q^2 - 0.2583 \cdot Q + 7.6220, \quad 0 \le Q \le 40, \quad \text{MNB} - 2, \text{Flat Speech} \tag{4}$$

$$AD \approx 0.0024 \cdot Q^2 - 0.2719 \cdot Q + 8.2523, \quad 0 \le Q \le 40, \quad \text{MNB} - 1, \text{IRS Filtered Speech} \tag{5}$$

$$AD \approx 0.0031 \cdot Q^2 - 0.2846 \cdot Q + 6.9276, \quad 0 \le Q \le 40, \quad \text{MNB} - 2, \quad \text{IRS Filtered Speech} \tag{6}$$

When coupled with (2), these results allow one to relate Q to L(AD). These relationships in turn allow reference to subjective test results that are given in terms of Q.

V. Conclusion

MNB structures estimate perceived speech quality by decomposing a codec output signal in a space defined partly by human hearing and judgment, and partly by the codec input signal. Nine ACR subjective tests, using the MOS scale were available for testing objective estimators of perceived speech quality. Together, these nine tests include 219 4-kHz bandwidth speech codecs, transmission systems, and reference conditions, with bit rates ranging from 2.4 to 64 kb/s. This collection of speech files and scores has allowed us to complete one of the most comprehensive tests of objective estimators of perceived relative speech quality. The new MNB-based estimators and six established estimators were tested.

These tests show that classical objective estimators based on SNR do not, in general, provide useful estimates of perceived speech quality. Pearson correlations between those estimates and subjective test results range from 0.22 to 0.64. More advanced methods like cepstral distance (correlations from 0.49 to 0.98) and Bark spectral distortion (correlations from 0.37 to 0.92) sometimes give high correlations but these methods still lack reliability when the broadest class of conditions is considered. The PSQM algorithm gives quite high correlations (0.98 to 0.99) for higher bit rate speech codecs operating over error-free channels, and impressive correlations over all (0.79 to 0.99). The MNB estimators provide the best results across all conditions considered in these tests.

When the MNB estimators were optimized using only two of the tests, they generalized well to the other seven tests. The correlations between subjective scores and the fully optimized MNB estimators range from .910 to .986. Given the breadth of conditions covered by the nine tests, these are very encouraging results. In particular, the improved ability to estimate perceived speech quality for lower rate speech codecs, some of which are operating with bit errors or frame erasures, represents an important advance. The two MNB structures presented and evaluated here were chosen for their balance of relatively low complexity and high performance as estimators of perceived speech quality across a wide range of conditions and quality levels. Other MNB structures may be more appropriate for more specific speech or

audio quality estimation applications. In addition, these structures or other MNB structures may address open issues in perceived audio quality estimation, layered speech or audio coding, automatic speech or speaker recognition, audio signal enhancement, and other areas.

Formal subjective tests will very likely always provide the final definitive word when codecs and transmission systems are evaluated in major standardization, marketing, and procurement decisions. But objective estimators of perceived relative speech quality have a role to play as well. That role continues to expand as new estimators, like those described here, demonstrate increased reliability across broader ranges of test conditions. Perceptually consistent objective estimators of speech quality can provide a meaningful common language for designers and developers who wish to compare their results, but do not have access to subjective testing facilities. Estimators may also be consulted to aid in design decisions that might otherwise be made on the basis of a single designer's perception and judgment. In this situation, a large number of talkers, languages, or other relevant conditions can be tested with little effort in a comparatively short time. Finally, objective estimators are particularly well-suited for continuously monitoring speech transmission and storage systems of interest, and reporting deviations from established baseline quality levels.

References

[1] CCITT Recommendation P.48, "Specification for an Intermediate Reference System," Geneva, 1989.

[2] N. Kitawaki, "Quality assessment of coded speech," in *Advances in Speech Signal Processing*, S. Furui and M. Sondi, Ed., New York: Marcel Dekker, 1992, pp. 357-385.

[3] Y. Be'ery, Z. Shpiro, T. Simchony, L. Shatz, and J. Piasetzky, "An efficient variable-bit-rate low-delay CELP coder," in *Advances in Speech Coding*, B.S. Atal, V. Cuperman, and A. Gersho, Eds., Boston: Kluwer Academic Publishers, 1990, pp. 37-46.

[4] S. Wang, A. Sekey, and A. Gersho, "An objective measure for predicting subjective quality of speech coders," *IEEE J. on Selected Areas in Communications*, vol. 10, pp. 819-829, Jun. 1992.

[5] ITU-T Recommendation P.861, "Objective Quality Measurement of Telephone-Band (300-3400 Hz) Speech Codecs," Geneva, 1996.

[6] J. G. Beerends and J. A. Stemerdink,"A perceptual speech-quality measure based on a psychoacoustic sound representation," *J. Audio Engineering Society*, vol. 42, pp. 115-123, Mar. 1994.

[7] S. Voran, "Techniques for comparing objective and subjective speech quality tests," in *Proc. Speech Quality Assessment Workshop at Ruhr-Universität Bochum, Germany*, 1994, pp. 59-64.

[8] ITU-T Recommendation P.861, Appendix II, "Objective Quality Measurement of Telephone-Band (300-3400 Hz) Speech Codecs Using Measuring Normalizing Blocks (MNBs)," Geneva, 1998.

[9] CCITT Recommendation P.81, "Modulated Noise Reference Unit (MNRU)," Geneva, 1989.

[10] A.V. McCree, et. al., "A 2.4 kbits/s MELP coder candidate for the new U.S. federal standard," in *Proc. IEEE ICASSP `96*, Atlanta, USA, May 1996, pp. 200-203.

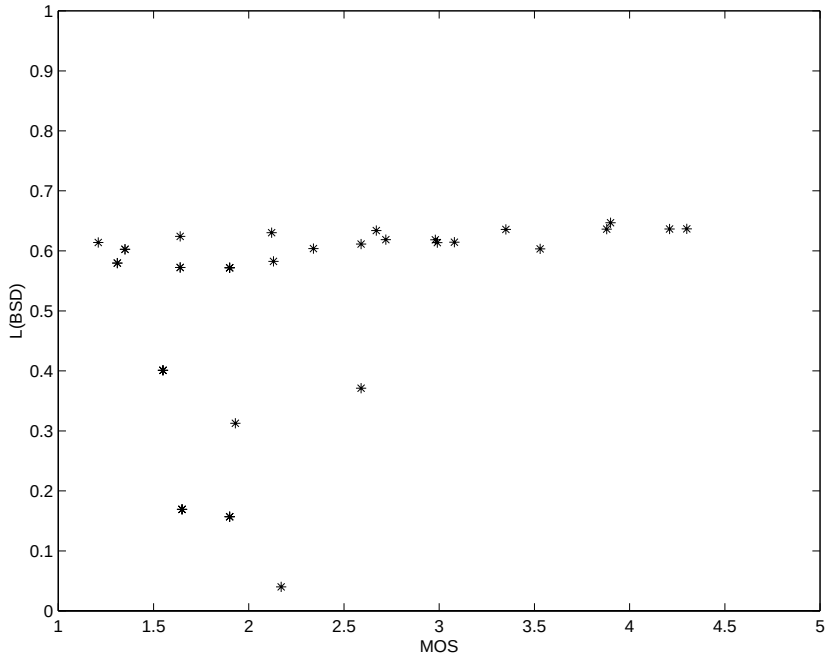

Figure 1. L(BSD) as an estimator of perceived speech quality on test 3, ρ=0.368.

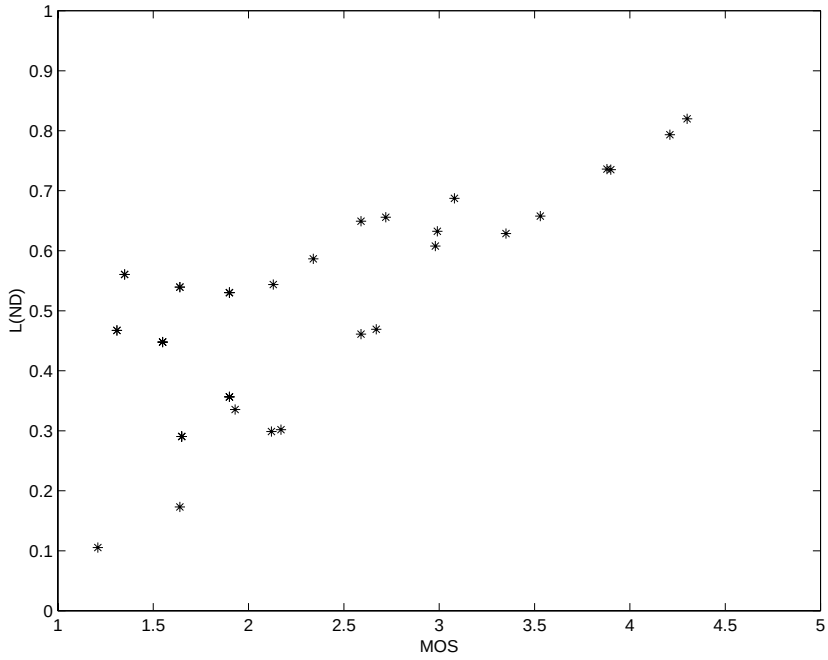

Figure 2. L(ND) as an estimator of perceived speech quality on test 3, ρ=0.793.

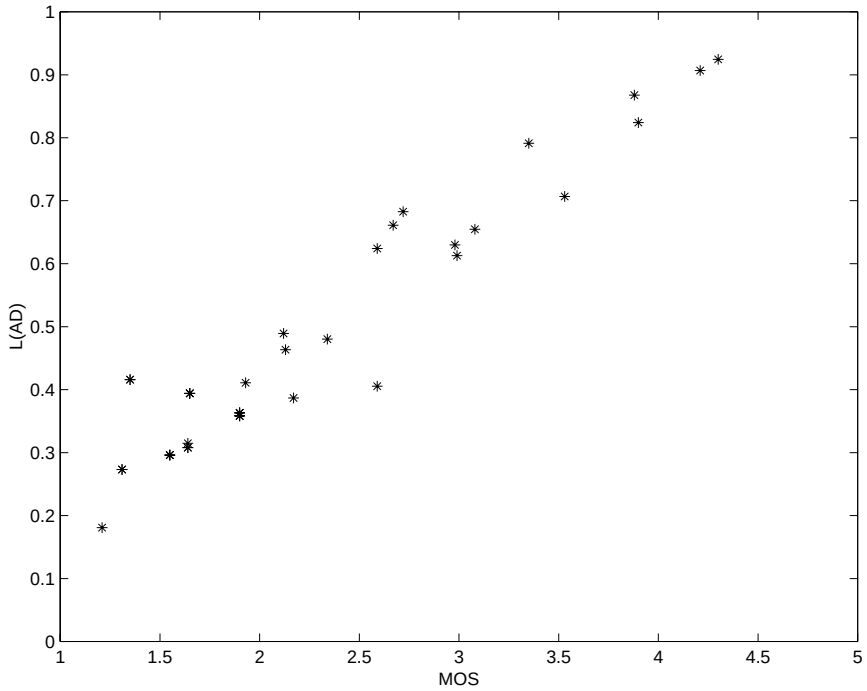

Figure 3. MNB structure 2 as an estimator of perceived speech quality on test 3, ρ=0.959.

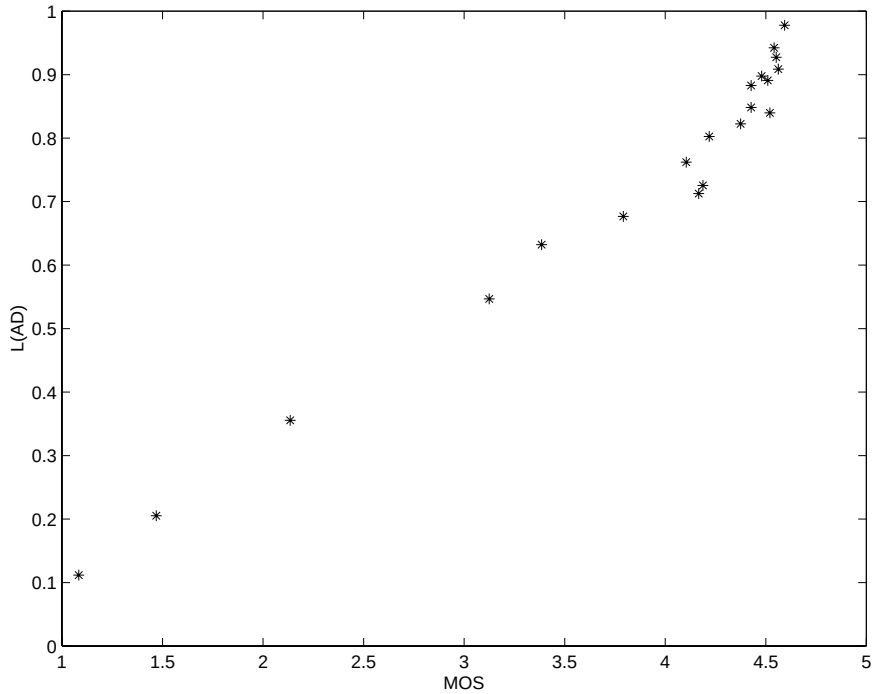

Figure 4. MNB structure 1 as an estimator of perceived speech quality on test 5, ρ=0.986.

TABLE I
SUMMARY OF MATERIAL IN NINE SUBJECTIVE TESTS

Test	Number of Conditions	Conditions [1,2]	Filtering of Input Speech	Language	Talkers per Condition	Files	Total Minutes
1	22	PCM: 64, 48, 40 kb/s ADPCM: 32 kb/s, x1, 2, 3, 4 APC: 16 kb/s, 2 versions Proprietary Codec: 16 kbps SELP: 4.8 kb/s, 2 versions LPC: 2.4 kb/s MNRU: 6 levels Narrow-Band MNRU: 3 levels	None	North American English	4	176	8
2	35	PCM: 64 kb/s Proprietary CELP A: 8 kb/s, over 9 RF channels, bit errors and frame erasures Proprietary CELP B: 8 kb/s, over 9 RF channels, bit errors and frame erasures AMPS over 9 RF channels MNRU: 7 levels	IRS filtered	North American English	6	1050	100
3	27	ADPCM: 32 kb/s, clear and bit errors CVSD: 32, 16 kb/s, clear and bit errors VSELP: 8 kb/s CELP: 4.8 kb/s, clear and bit errors IMBE: 4.8, 2.4 kb/s STC A: 4.8, 2.4 kb/s, clear and bit errors STC B: 2.4 kb/s LPC: 2.4 kb/s, clear and bit errors POTS MNRU: 8 levels	None	North American English	6	1994	225

4	38	ADPCM: 32 kb/s, x4 LD-CELP: 16 kb/s VSELP: 8 kb/s Proprietary Non-Waveform Codec: 6.4 kb/s Proprietary Non-Waveform Codec: 4 kb/s, 3 input levels Proprietary Non-Waveform Codec: 4 kb/s, x2 Proprietary Non-Waveform Codec: 4 kb/s + ADPCM: 32 kb/s Proprietary Non-Waveform Codec: 4 kb/s + VSELP: 8 kb/s Proprietary Non-Waveform Codec: 4 kb/s + RPE-LTP: 13 kb/s Proprietary Non-Waveform Codec: 4 kb/s + LD-CELP: 16 kb/s + LD-CELP: 16 kb/s MNRU: 7 levels	Both IRS filtered and unfiltered	North American English	8	2432	264
5	20	PCM: 64 kb/s, x1, 2, 4, 8, 16 ADPCM: 32 kb/s, x1, 2, 4 G.728 Candidate 16 kb/s, x1, 2, 4 MNRU: 9 levels	IRS filtered	North American English	4	1440	206
6	20	Same as test 5	IRS filtered	Japanese	4	1440	188
7	20	Same as test 5	IRS filtered	Italian	4	1440	131
8	47	LD-CELP: 16 kb/s 8 CELP Codecs: ≅13 kb/s, frame error rates 0, 1, 2, 3, 5% MNRU: 6 levels	IRS filtered	North American English	8	1360	136
9	30	VSELP: 8 kb/s, 11 simulated radio environments ACELP: 8 kb/s, 11 simulated radio environments PCM: 64 kb/s CELP: 4.8 kb/s POTS MNRU: 5 levels	Both IRS filtered and unfiltered	North American English	8	480	54

[1] The notation “xN” is used to indicate N passes through the indicated device.

[2] The notation “codec1 + codec2” is used to indicate that two different codecs were tandemed to create a single condition.

TABLE II
OPTIMIZED VALUES OF LOGISTIC FUNCTION PARAMETERS

Objective Estimator	*a*	*b*
SNR	-0.0552	-0.3490
SNRseg	-0.0542	-0.3927
PWSNRseg	-0.1073	0.1910
CD	0.4175	-1.8274
BSD	6.3081	-0.7434
ND	0.5567	-1.7450

TABLE III
PER-CONDITION PEARSON COEFFICIENTS OF CORRELATION BETWEEN SUBJECTIVE SCORES AND OBJECTIVE ESTIMATORS

Test	L(SNR)	L(SNRseg)	L(PWSNRseg)	L(CD)	L(BSD)	L(ND)
1*	.333	.381	.393	.486	.825	.928
2*	.526	.522	.620	.729	.731	.941
3*	.295	.494	.507	.617	.368	.793
4*	.247	.221	.636	.789	.863	.973
5	.226	.267	.523	.948	.919	.986
6	.271	.313	.502	.933	.850	.986
7	.317	.340	.542	.975	.892	.976
8*	.556	.381	.605	.671	.801	.858
9*	.433	.326	.544	.838	.712	.827

* These tests include conditions that are outside the defined scope of the PSQM (ND) algorithm.

TABLE IV
PER-CONDITION PEARSON COEFFICIENTS OF CORRELATION BETWEEN SUBJECTIVE SCORES AND OBJECTIVE ESTIMATORS

Test	L(ND)	L(AD)			
		MNB-1	MNB-2	MNB-1	MNB-2
		Weights optimized using only tests 1 and 2.		Weights optimized using tests 1-9.	
1	.928	.931	.928	.932	.956
2	.941	.965	.963	.951	.945
3	.793	.939	.944	.935	.959
4	.973	.964	.979	.977	.976
5	.986	.955	.963	.986	.984
6	.986	.965	.969	.983	.982
7	.976	.967	.971	.980	.984
8	.858	.954	.953	.936	.961
9	.827	.921	.923	.910	.942

TABLE V
AVERAGE SENSITIVITY OF MNB ESTIMATORS TO TEMPORAL MISALIGNMENT

Temporal Misalignment	Average Percentage Drop in L(AD)	
	Coarse Group	Fine Group
1 ms	0 %	10 %
2 ms	4 %	19 %
4 ms	8 %	34 %
8 ms	17 %	43 %
16 ms	48 %	67 %
32 ms	85 %	91 %

TABLE VI
MNB STRUCTURE 1 BENCHMARK VALUES FOR FLAT SPEECH

Condition	Mean AD	Half-width of 95% CI on Mean AD	Mean L(AD)	Half-width of 95% CI on Mean L(AD)
G.711 PCM, μ-law, 64 kbps	1.9144	0.0645	0.9395	0.0040
G.711 PCM, A-law, 64 kbps	1.8565	0.0631	0.9428	0.0036
G.726 ADPCM μ-law, 40 kbps	2.3810	0.0545	0.9077	0.0048
G.726 ADPCM A-law, 40 kbps	2.3517	0.0522	0.9103	0.0045
G.726 ADPCM μ-law, 32 kbps	2.9522	0.0543	0.8480	0.0070
G.726 ADPCM A-law, 32 kbps	2.9450	0.0538	0.8489	0.0069
G.726 ADPCM μ-law, 24 kbps	3.9458	0.0571	0.6753	0.0121
G.726 ADPCM A-law, 24 kbps	3.9257	0.0615	0.6793	0.0130
G.726 ADPCM μ-law, 16 kbps	5.1584	0.0745	0.3866	0.0176
G.726 ADPCM A-law, 16 kbps	5.1490	0.0779	0.3890	0.0183
G.728 LD-CELP, 16 kbps	3.2460	0.0710	0.8048	0.0112
GSM 6.10 RPE-LTP, 13 kbps	3.3194	0.0532	0.7949	0.0086
TIA/EIA 635 VSELP, 8 kbps	3.5978	0.0531	0.7462	0.0100
FS1016 CELP, 4.8 kbps	4.2856	0.0532	0.5981	0.0127
FS1015 LPC, 2.4 kbps	4.9589	0.0684	0.4340	0.0164
MELP, 2.4 kbps [10]	4.4928	0.0748	0.5475	0.0182
MNRU, Q=40	1.5366	0.0365	0.9586	0.0015
MNRU, Q=36	1.8960	0.0522	0.9411	0.0030
MNRU, Q=35	2.0097	0.0568	0.9343	0.0036
MNRU, Q=30	2.7244	0.0785	0.8728	0.0086
MNRU, Q=25	3.6246	0.0933	0.7368	0.0171
MNRU, Q=24	3.8173	0.0951	0.6986	0.0189
MNRU, Q=20	4.6089	0.1020	0.5182	0.0244
MNRU, Q=18	5.0027	0.1059	0.4244	0.0252
MNRU, Q=15	5.5805	0.1127	0.2985	0.0236
MNRU, Q=12	6.1346	0.1209	0.2013	0.0198
MNRU, Q=10	6.4870	0.1272	0.1532	0.0169
MNRU, Q=6	7.1354	0.1388	0.0893	0.0115
MNRU, Q=5	7.2862	0.1414	0.0783	0.0103
MNRU, Q=0	7.9791	0.1497	0.0418	0.0059

TABLE VII
MNB STRUCTURE 2 BENCHMARK VALUES FOR FLAT SPEECH

Condition	Mean AD	Half-width of 95% CI on Mean AD	Mean L(AD)	Half-width of 95% CI on Mean L(AD)
G.711 PCM, μ-law, 64 kbps	0.8605	0.0334	0.8997	0.0030
G.711 PCM, A-law, 64 kbps	0.8251	0.0320	0.9029	0.0028
G.726 ADPCM μ-law, 40 kbps	1.1822	0.0296	0.8669	0.0034
G.726 ADPCM A-law, 40 kbps	1.1636	0.0296	0.8690	0.0033
G.726 ADPCM μ-law, 32 kbps	1.6170	0.0406	0.8078	0.0063
G.726 ADPCM A-law, 32 kbps	1.6126	0.0382	0.8087	0.0059
G.726 ADPCM μ-law, 24 kbps	2.4503	0.0545	0.6465	0.0124
G.726 ADPCM A-law, 24 kbps	2.4341	0.0598	0.6499	0.0135
G.726 ADPCM μ-law, 16 kbps	3.6229	0.0824	0.3665	0.0187
G.726 ADPCM A-law, 16 kbps	3.6280	0.0877	0.3660	0.0196
G.728 LD-CELP, 16 kbps	1.8195	0.0454	0.7743	0.0080
GSM 6.10 RPE-LTP, 13 kbps	1.6594	0.0419	0.8011	0.0066
TIA/EIA 635 VSELP, 8 kbps	2.1782	0.0461	0.7060	0.0095
FS1016 CELP, 4.8 kbps	2.7902	0.0486	0.5667	0.0118
FS1015 LPC, 2.4 kbps	3.8886	0.0790	0.3084	0.0163
MELP, 2.4 kbps [10]	3.0911	0.0959	0.4935	0.0232
MNRU, Q=40	0.6219	0.0214	0.9196	0.0016
MNRU, Q=36	0.8669	0.0324	0.8991	0.0029
MNRU, Q=35	0.9468	0.0359	0.8915	0.0034
MNRU, Q=30	1.4778	0.0554	0.8274	0.0076
MNRU, Q=25	2.2351	0.0770	0.6915	0.0155
MNRU, Q=24	2.4129	0.0818	0.6527	0.0175
MNRU, Q=20	3.1958	0.1017	0.4669	0.0243
MNRU, Q=18	3.6213	0.1123	0.3686	0.0255
MNRU, Q=15	4.2878	0.1272	0.2382	0.0237
MNRU, Q=12	4.9660	0.1402	0.1428	0.0187
MNRU, Q=10	5.4123	0.1475	0.0991	0.0149
MNRU, Q=6	6.2511	0.1596	0.0476	0.0084
MNRU, Q=5	6.4478	0.1624	0.0398	0.0072
MNRU, Q=0	7.3357	0.1727	0.0173	0.0033

TABLE VIII
MNB STRUCTURE 1 BENCHMARK VALUES FOR IRS FILTERED SPEECH

Condition	Mean AD	Half-width of 95% CI on Mean AD	Mean L(AD)	Half-width of 95% CI on Mean L(AD)
G.711 PCM, μ-law, 64 kbps	1.6095	0.0406	0.9554	0.0019
G.711 PCM, A-law, 64 kbps	1.5766	0.0390	0.9569	0.0017
G.726 ADPCM μ-law, 40 kbps	2.6178	0.0504	0.8863	0.0052
G.726 ADPCM A-law, 40 kbps	2.6055	0.0493	0.8876	0.0050
G.726 ADPCM μ-law, 32 kbps	3.2749	0.0554	0.8018	0.0088
G.726 ADPCM A-law, 32 kbps	3.2733	0.0542	0.8022	0.0086
G.726 ADPCM μ-law, 24 kbps	4.1863	0.0537	0.6214	0.0125
G.726 ADPCM A-law, 24 kbps	4.1845	0.0542	0.6219	0.0127
G.726 ADPCM μ-law, 16 kbps	5.3573	0.0688	0.3413	0.0153
G.726 ADPCM A-law, 16 kbps	5.3607	0.0686	0.3405	0.0152
G.728 LD-CELP, 16 kbps	3.2370	0.0630	0.8070	0.0101
GSM 6.10 RPE-LTP, 13 kbps	3.6603	0.0582	0.7339	0.0112
TIA/EIA 635 VSELP, 8 kbps	3.8011	0.0700	0.7049	0.0145
FS1016 CELP, 4.8 kbps	4.3568	0.0716	0.5803	0.0170
FS1015 LPC, 2.4 kbps	5.2181	0.0911	0.3743	0.0205
MELP, 2.4 kbps [10]	4.8443	0.0701	0.4614	0.0171
MNRU, Q=40	1.4121	0.0288	0.9634	0.0011
MNRU, Q=36	1.6163	0.0393	0.9552	0.0018
MNRU, Q=35	1.6834	0.0431	0.9521	0.0021
MNRU, Q=30	2.1452	0.0667	0.9248	0.0052
MNRU, Q=25	2.8320	0.0952	0.8583	0.0127
MNRU, Q=24	2.9971	0.1001	0.8369	0.0147
MNRU, Q=20	3.7362	0.1180	0.7124	0.0246
MNRU, Q=18	4.1487	0.1237	0.6255	0.0285
MNRU, Q=15	4.8046	0.1278	0.4736	0.0301
MNRU, Q=12	5.4782	0.1285	0.3226	0.0261
MNRU, Q=10	5.9331	0.1279	0.2357	0.0216
MNRU, Q=6	6.8134	0.1248	0.1160	0.0124
MNRU, Q=5	7.0248	0.1239	0.0963	0.0105
MNRU, Q=0	7.9904	0.1221	0.0395	0.0047

TABLE IX
MNB STRUCTURE 2 BENCHMARK VALUES FOR IRS FILTERED SPEECH

Condition	Mean AD	Half-width of 95% CI on Mean AD	Mean L(AD)	Half-width of 95% CI on Mean L(AD)
G.711 PCM, μ-law, 64 kbps	0.7007	0.0230	0.9135	0.0018
G.711 PCM, A-law, 64 kbps	0.6783	0.0219	0.9153	0.0017
G.726 ADPCM μ-law, 40 kbps	1.4589	0.0433	0.8309	0.0062
G.726 ADPCM A-law, 40 kbps	1.4513	0.0436	0.8320	0.0062
G.726 ADPCM μ-law, 32 kbps	2.0275	0.0560	0.7354	0.0112
G.726 ADPCM A-law, 32 kbps	2.0159	0.0543	0.7378	0.0109
G.726 ADPCM μ-law, 24 kbps	2.8975	0.0739	0.5405	0.0180
G.726 ADPCM A-law, 24 kbps	2.8905	0.0714	0.5421	0.0175
G.726 ADPCM μ-law, 16 kbps	3.9852	0.0940	0.2904	0.0181
G.726 ADPCM A-law, 16 kbps	3.9820	0.0937	0.2911	0.0179
G.728 LD-CELP, 16 kbps	1.9666	0.0455	0.7477	0.0085
GSM 6.10 RPE-LTP, 13 kbps	1.9071	0.0454	0.7587	0.0083
TIA/EIA 635 VSELP, 8 kbps	2.4007	0.0620	0.6572	0.0139
FS1016 CELP, 4.8 kbps	2.8412	0.0687	0.5536	0.0166
FS1015 LPC, 2.4 kbps	4.1366	0.1037	0.2622	0.0188
MELP, 2.4 kbps [10]	3.4433	0.0863	0.4085	0.0201
MNRU, Q=40	0.5631	0.0201	0.9238	0.0015
MNRU, Q=36	0.7183	0.0268	0.9120	0.0022
MNRU, Q=35	0.7698	0.0291	0.9077	0.0025
MNRU, Q=30	1.1213	0.0450	0.8730	0.0052
MNRU, Q=25	1.6646	0.0667	0.7982	0.0110
MNRU, Q=24	1.7990	0.0710	0.7755	0.0126
MNRU, Q=20	2.4236	0.0899	0.6500	0.0203
MNRU, Q=18	2.7858	0.0978	0.5662	0.0235
MNRU, Q=15	3.3875	0.1084	0.4230	0.0254
MNRU, Q=12	4.0407	0.1176	0.2828	0.0226
MNRU, Q=10	4.4979	0.1226	0.2034	0.0188
MNRU, Q=6	5.4260	0.1309	0.0948	0.0105
MNRU, Q=5	5.6576	0.1326	0.0772	0.0089
MNRU, Q=0	6.7363	0.1414	0.0285	0.0038