INTERNATIONAL TELECOMMUNICATION UNION

Additions to the Handbook on Telephonometry

ITU-T

ITU-T
TELECOMMUNICATION
STANDARDIZATION
SECTOR OF ITU

1999

INTERNATIONAL TELECOMMUNICATION UNION

Additions to the Handbook on Telephonometry

ITU-T
TELECOMMUNICATION
STANDARDIZATION
SECTOR OF ITU

1999

ADDITIONS TO THE HANDBOOK ON TELEPHONOMETRY

CONTENTS

	Page
Addition to § 2.3.6 of the Handbook on Telephonometry	1
Addition to § 3.2.5 of the Handbook on Telephonometry	7
Addition to § 4.8 of the Handbook on Telephonometry	9

Addition to § 2.3.6 of the Handbook on Telephonometry

SOME EFFECTS OF SIDETONE

1 Introduction

Over a number of years, sidetone has been studied and some important conclusions have been reached from the point of view of the subscriber in his role as both talker and listener. These conclusions relate to the effect of sidetone on a subscriber as he hears his own voice, the way his talking level changes as a result and some effects of sidetone when the subscriber is listening in conditions of moderate to high-level room noise. These effects are summarized in Figures 1 and 3.

The relationship between talker and listener sidetone for a given telephone depends primarily on two factors:

a) the geometry of its handset; and

b) whether or not there are any non-linear gain or loss characteristics in the sidetone path.

Some guidance for telephone set designers is provided in clause 4.

Some information is also provided concerning the increasingly frequent occurrence of short delay talker echo, which may be perceived as unpleasant talker sidetone.

2 Talker sidetone

Figure 1 shows that there is a preferred range for sidetone when the subscriber is talking under quiet conditions, and that the difference between the sidetone being objectionable or too quiet is of the order of 20 dB. (These results were obtained from talking-only tests and need to be confirmed by conversation tests.) The preferred range lies between 7 and 12 dB, STMR (Sidetone Masking Rating – Recommendation P.76 [1]).

The acceptable range is wider and lies between an STMR of 1 dB and 17 dB (although it must be stated that increasing STMR to a value greater than 17 dB is likely to affect only the talking level, and that only marginally). This range corresponds to the difference between the two curves at the 50% appraisals level. It is not proposed that the 17 dB figure should in any way be considered a maximum value. However, for an STMR above 20 dB, the connection sounds "dead".

Figure 2 shows the way in which the talking level changes with sidetone level. These results were obtained by means of conversation tests, for a connection close to the preferred overall loss. For telephone connections where the OLR is in the preferred range, the STMR values may be positioned in the preferred STMR range given above. However, on high-loss connections the STMR value should be close to, or even exceed, 12 dB to encourage the subscriber to speak louder. On low-loss connections, the STMR value may be sometimes permitted to become less than 7 dB, but only rarely should it become as low as 1 dB, e.g. telephone sets with receive volume control. Recommendation G.121 [2] interprets those results for transmission planning purposes.

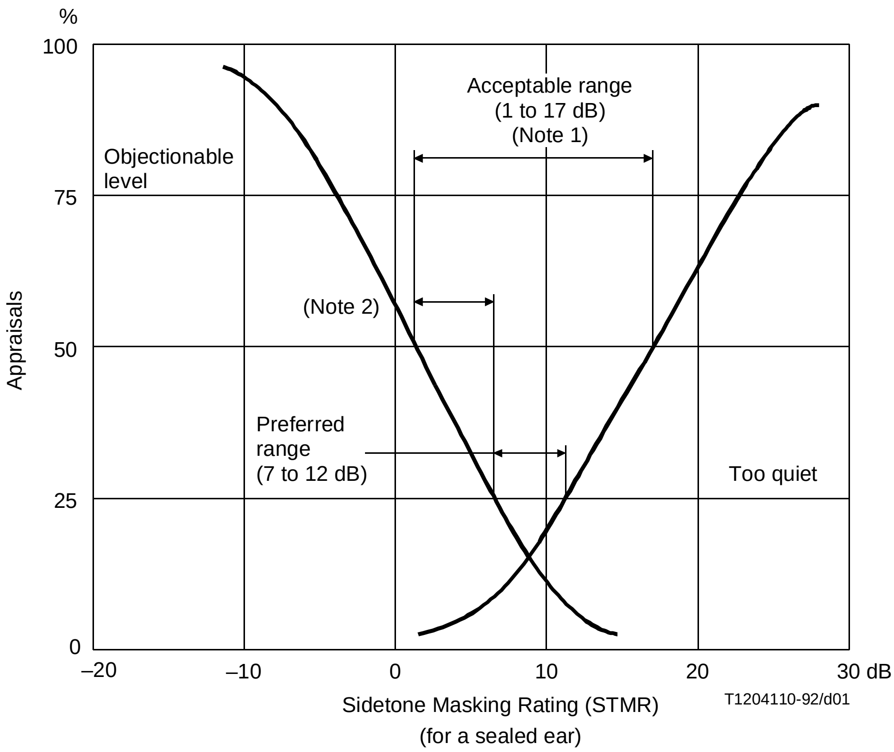
The speech voltage will also be a function of room noise level. For modern telephones with linear transmitters, it may be desirable to have the STMR value in the range from 10 to 15 dB if it is expected that telephones will be used in noisy environments.

3 Listener sidetone

High room noise in the subscriber's environment disturbs the received speech in two ways:

i) noise being picked up by the handset microphone and transmitted to the handset receiver via the electric sidetone path;

ii) noise leaking past the earcap at the handset receiver.



NOTE 1 – Conversational conditions will determine what part of this range is acceptable for a given connection.

NOTE 2 – This part of the acceptable range (1 to 7 dB) should only be entered with caution, e.g. on low-loss connections (see Recommendation G.121) or where there is a receive volume control.

Figure 1 – Curves showing sidetone levels that are objectionable and too quiet, together with the preferred range, for the subscriber as a talker

Studies have shown that at low frequencies the earcap leakage path dominates over the electric sidetone path in much the same way as the human sidetone signal does in talker sidetone. The weightings applied in the STMR loudness calculation are therefore applicable and the Listener Sidetone Rating (LSTR, Recommendation P.76) has been developed, which makes use of the room noise sidetone sensitivity (see clause 9/P.64 [3]) in the STMR rating method (see Recommendation P.79 [4]).

Results of subjective tests of LSTR vs MOS (using in this case a mean opinion scale of 0-10) are given in Figure 3. In each case, the LSTR was derived by making use of Δ_{Sm} (see Recommendations P.10 [5], P.64, P.79 and 3.3.17, Part C in the Handbook on Telephonometry [6]) to convert the sidetone sensitivities S_{meST} to S_{RNST} before calculating LSTR or applied as a weighted correction to STMR as described in A.4.3.3/G.111 [7]. Room noise levels were comparable at 55-59 dBA.

Based upon these results, Recommendation G.121 recommends that a value of 13 dB LSTR should be striven for.

The value 13 dB is based on a 10 dB LSTR (which may be considered a minimum value), where no further improvement in mean opinion score was possible by increasing LSTR (see Figure 3), plus an allowance of 3 dB reflecting the fact that room noise in some office locations can exceed the values used in these experiments. Other tests have also suggested that a higher figure might be more appropriate.

The value that is satisfactory in a given telephone connection will depend on such factors as the level of room noise, the OLR of the connection, the talking levels used, etc. In particular, modern telephones with linear transmitters are more efficient at picking up background room noise. In this case, it may be desirable to have LSTR > 15 dB.

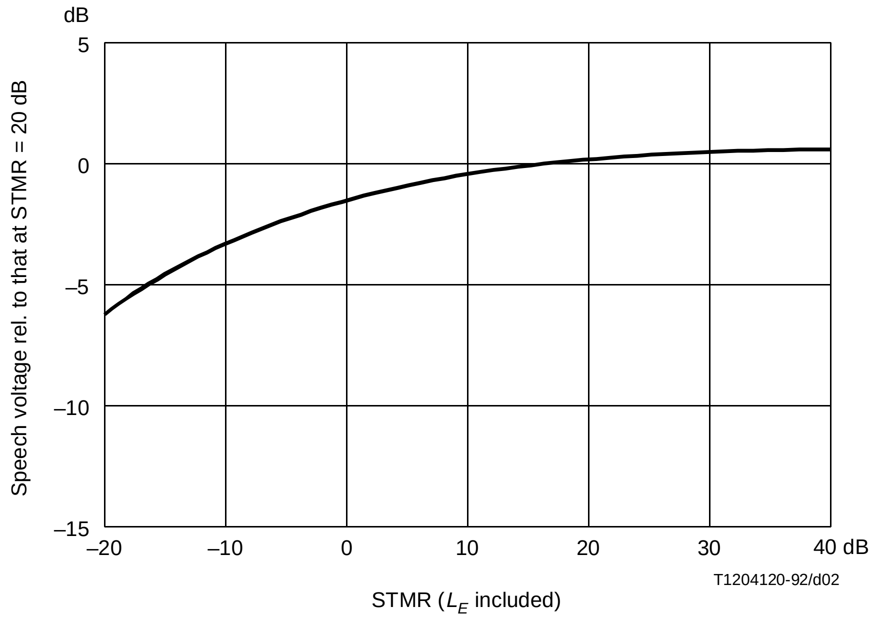


Figure 2 – Speech voltage as a function of STMR

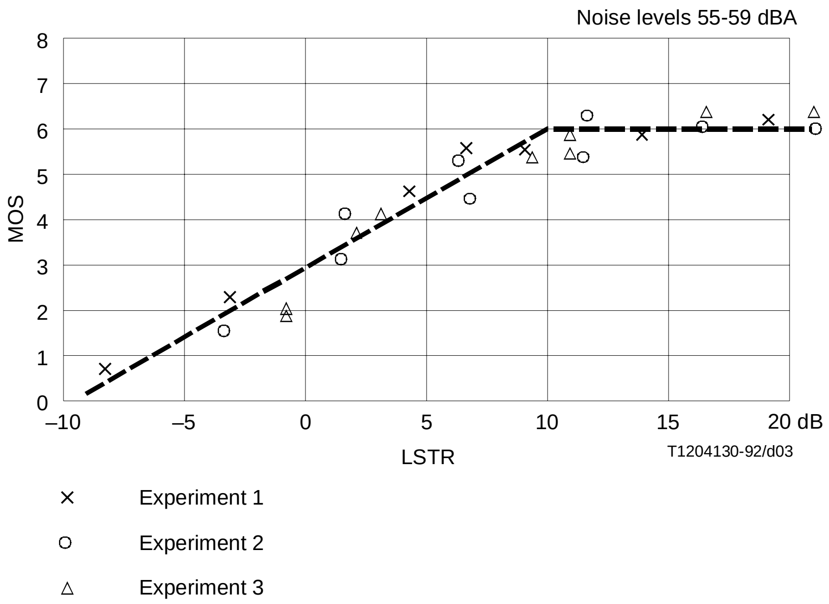


Figure 3 – MOS as a function of LSTR calculated from different test results

4 Relationship between talker and listener sidetone

4.1 Telephones having linear sidetone characteristics

For telephones having linear gain or loss characteristics in the sidetone path, the relationship between talker and listener sidetone levels is controlled by the geometry of the handset. There are two aspects of the geometry that seem most important: the distance from the mouth to the transmitter port and the size of the obstacle created by the transmitter end.

For speech inputs, a handset having a large transmitter end positioned close to the mouth experiences a greater sound pressure at its transmitter port than a handset having the transmitter end positioned farther from the mouth (distance effect) or one having a small transmitter end (obstacle effect). However, for diffuse field room noise inputs, the sound pressure at the transmitter port is independent of the size and shape of the handset. Thus, if the STMR level is the same for the two handsets, the one with the large transmitter end close to the mouth will have less electrical gain in its sidetone path, which will result in a greater LSTR value.

It has been shown that the difference in LSTR and STMR levels for a sample of 26 linear telephone sets is highly correlated to the logarithm of the distance between the transmitter port (centre of the external opening for the microphone on the surface of the handset) and the centre of the lip ring of the artificial mouth when the handset is placed in the LRGP test position (see Recommendation P.64). It has the following empirical relationship:

$$LSTR - STMR = 33 - 20 \log(d)$$

where the distance d from the transmitter port to the centre of the lip ring is measured in millimetres. There may be small perturbations in the order of ± 1 dB about this relationship depending on the obstacle size presented by the transmitter end of the handset.

NOTE – This relationship is based on measurements of telephones having more or less conventional handsets. It may not be applicable for handsets with extreme geometries or for operator headsets that have their transmitter ports located behind the lip plane.

4.2 Telephones having non-linear sidetone characteristics

Non-linear gain or loss characteristics may be used in the electrical sidetone path to increase the LSTR – STMR difference. Carbon transmitters, for example, frequently are less sensitive to the lower input levels of room noise than they are to the higher input levels of speech. Such a characteristic may be introduced into telephones having linear microphones through the use of various non-linear gain circuits.

If the same non-linear gain function is used in both the send and sidetone paths of the telephone, then the LSTR – STMR difference may be approximated by measuring the difference in send sensitivities due to speech and room noise inputs, DELSM, as described in Recommendation P.64. A STMR difference may then be calculated according to the method given in Annex A/G.111. However, if the send and sidetone paths do not have the same non-linear gain characteristics (e.g. automatic gain control circuitry in the receive path that affects sidetone), then the DELSM method will give erroneous results. In this case, the LSTR and STMR values must be measured directly.

5 Short delay talker echo perceived as sidetone

Talker echo can have a detrimental effect on transmission quality at delay times of a few milliseconds, even though the delay is not long enough for it to be perceived as an echo signal separate from the sidetone. Such echoes can occur, for example, due to reflections from the analogue trunk port of a digital PBX or on local analogue calls through a digital exchange. Unless the hybrid that converts the 4-wire digital PBX or exchange back to a 2-wire analogue circuit is well matched, some reflections will occur. Because of the digital processing times involved, these talker echo signals have a few milliseconds of delay. Sidetone provides a beneficial masking of low-level short delay talker echo, but as the talker echo level increases, it interacts with the sidetone in an unpleasant manner (hollow-sounding sidetone, rain-barrel effect, etc.).

The objectively measurable effect of short delay talker echo is that it produces ripples in the sidetone frequency response. The reflected talker echo signal is added to the direct sidetone signal with a phase relationship that increases the signal at some frequencies and decreases it at others. The spacing between the ripples is equal to the reciprocal of the delay. When the reflected talker echo signal is small relative to the direct sidetone, the ripples are small. As the talker echo signal increases in magnitude, the ripples increase in size until the peaks are 6 dB above the in-phase signal and the troughs are very deep due to almost exact out-of-phase cancellation. At even higher talker echo levels (or lower sidetone levels) the amount of ripple again decreases, but the predominant signal is then the delayed talker echo.

People perceive short delay talker echo combined with sidetone differently than an equivalent level of pure sidetone, even though they may not be able to detect that a separate echo signal is present. Thus, a simple sidetone measure such as STMR is not adequate to describe the effect of the combined signal. Talker echo, even with very short delay times, must be treated as a separate impairment to transmission quality. Recommendation P.11 [8] and reference [9] provide some guidance as to how both sidetone and talker echo may be taken into account in predicting the quality of a telephone connection, but this subject remains under study.

References

[1] CCITT Recommendation P.76 (1988), *Determination of loudness ratings; fundamental principles*.

[2] ITU-T Recommendation G.121 (1993), *Loudness Ratings (LRs) of national systems*.

[3] ITU-T Recommendation P.64 (1997), *Determination of sensitivity/frequency characteristics of local telephone systems*.

[4] ITU-T Recommendation P.79 (1993), *Calculation of loudness ratings for telephone sets*.

[5] ITU-T Recommendation P.10 (1998), *Vocabulary of terms on telephone transmission quality and telephone sets*.

[6] *Handbook on telephonometry*, ITU, Geneva, 1993.

[7] ITU-T Recommendation G.111 (1993), *Loudness ratings (LRs) in an international connection*.

[8] ITU-T Recommendation P.11 (1993), *Effect of transmission impairments*.

[9] Addition to § 3.2.5 of the Handbook on Telephonometry: *Artificial conversational speech,* ITU, Geneva, 1999.

Addition to § 3.2.5 of the Handbook on Telephonometry

ARTIFICIAL CONVERSATIONAL SPEECH

The Artificial Voice described in Recommendation P.50 is used as a one-way test signal in the performance evaluation of, for example, low bit-rate speech codecs. It is required that the application area of these artificial signals be enlarged for the performance evaluation of devices that are operated by speech signals, such as echo cancellers and voice switches in loudspeaker telephone sets and DSI (Digital Speech Interpolation) devices. For these purposes, artificial signals should simulate conversations by humans. Therefore, they should contain not only "talkspurt" periods expressed in terms of the Artificial Voices in Recommendation P.50 but also "pause" periods. Artificial signals are also required to be two-way signals for simulating a "talker and listener" environment.

A signal that alternately places the Artificial Voice in Recommendation P.50 of arbitrary duration and silence (zero sequences) of arbitrary duration along a time axis is not adequate, because human conversational speech has specific characteristics as conversation. For example, humans cannot continue to utter for a long period without pause and there are few cases of two persons uttering simultaneously for a long period.

Therefore, some statistical temporal characteristics of real conversational speech should be simulated in artificial signals. An artificial signal that meets this condition is called "Artificial Conversational Speech".

Figure 1 illustrates temporal power patterns when two persons talk to each other. In Artificial Conversational Speech, the following correspond well to the characteristics of real conversational speech: such statistical characteristics of talkspurt, pause, double talk and mutual silence as occurrence rates, average values and cumulative distribution of duration.

Figure 2 shows a state transition model for Artificial Conversational Speech generation. A conversation is classified into four states: single talk (one talks and the other is silent, and vice versa), double talk, and mutual silence; and a state transition model is introduced among these four states. After staying at one state for a duration T, it transits from one state to another by the transition probability pi. The cumulative distributions of duration T in each state are assumed to be exponential and the duration T of each state varies according to random variable x. In silence periods, zero sequences are output. In talk periods, the Artificial Voice in Recommendation P.50 is output.

Parameters to be optimized in Figure 2 are as follows: estimated average duration in single talk, double talk, and mutual silence; and transition probability p1. Transition probabilities p2 and p3 are fixed at 50%. These parameters were optimized to minimize the average deviation of the following in artificial signals from those in real conversational speech: the average duration and the rate of talkspurt, pause, double talk and mutual silence. For the target values for parameter optimization, the average values measured for Japanese, American English, and Italian conversational speech signals were used. The target values for parameter optimization and the optimum parameter values are specified in Recommendation P.59.

The statistical temporal characteristics in the Artificial Conversational Speech tend to become stable as the total duration of the signals increases. The Artificial Conversational Speech exhibits a convergence time of long-term characteristics of ten minutes, which corresponds to the long-term characteristics of human conversational speech of about 10 hours. That shows the effectiveness of the use of the Artificial Conversational Speech instead of human conversational speech for measurements.

Artificial Conversational Speech is not normally generated by implementing the P.59 process in real time, but is generally stored either in ROM or on magnetic or optical media. In both cases, calibration facilities are usually provided. These consist of trailing tones that have a known relationship to the level of Artificial Conversational Speech.

Procedure

Artificial Conversational Speech requires no calibration. It needs only level adjustment by using the trailing tones available on the digital or analogue support where it is stored (ROM, CD, DAT, MO, PCM video recordings, or analogue tapes).

Artificial Conversational Speech can be used either as an electric signal or may be fed to the Artificial Mouth to obtain an Acoustical Artificial Conversational Signal. In the latter case, the Artificial Mouth must be provided with an equalizer. The correct mouth equalizer is checked by measuring the average spectrum of the Acoustical Artificial Speech Signal at MRP.

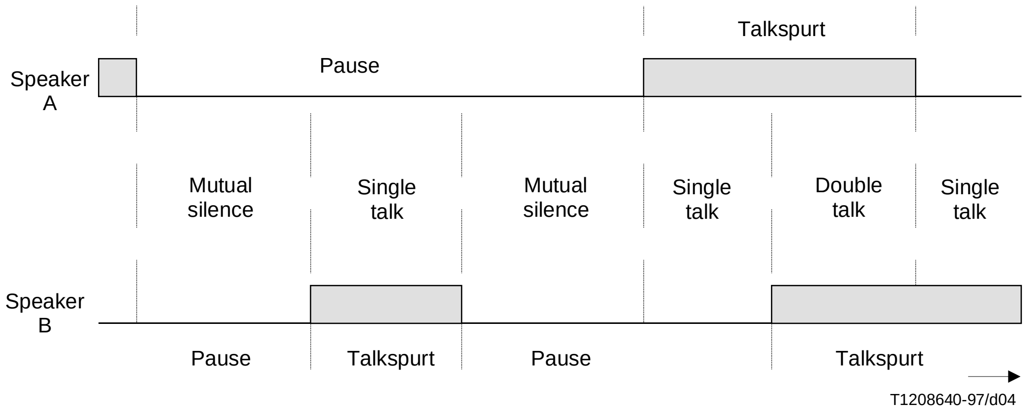
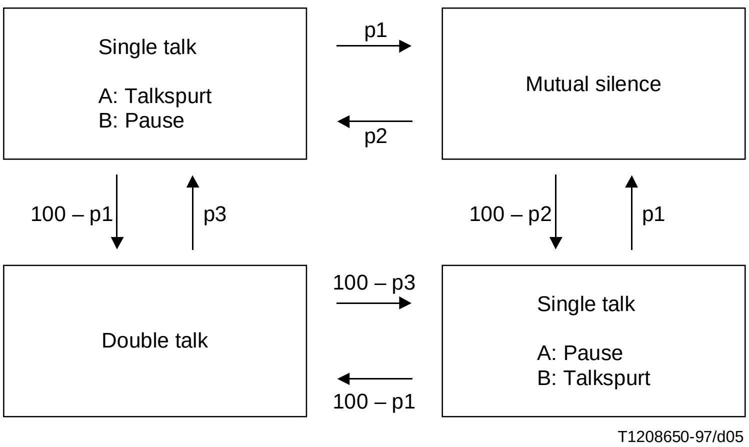


Figure 1 – Temporal power pattern in conversation



pi	Transition probability
T	Duration in each state
$\hat{T}$	Estimated average duration
$0 < x < 1$	Random variable

$T = -\hat{T}\,\ell n\,(1 - x)$

Figure 2 – State transition model for conversation

COMPUTATION MODELS FOR ESTIMATING CUSTOMER OPINIONS ABOUT SPEECH COMMUNICATION QUALITY IN TELEPHONE NETWORKS

1 Introduction

It is important to achieve a high probability that users are provided with adequate transmission quality for voice telephony. The traditional transmission planning methods – allocating permissible ranges for the individual transmission parameters – are being found as not fully adequate for modern networks. Among the reasons for this are the introduction of new technologies, for instance low-rate codecs, and the on-going de-regulation of the telecom market.

In these changing circumstances, network operators need guidance and a satisfactory method for assessing, already in the planning stage, the transmission quality of connections and predicting whether or not the user will be satisfied. The methodology needs to be based on computation and planning values, because it is not possible for measurements to be made to check the performance of all possible paths through real or simulated networks.

This has created a renewed interest in opinion estimations by computation models which can handle the combination effects of different types of transmission impairments. In what follows, a short survey will be given of such models which have been used in the ITU-T technical domain as a help for transmission planning. However, before the different computation models are presented, the concept "customer opinion" will be discussed in order to avoid misunderstandings about subjective tests and customer surveys.

2 "Customer opinions" as a basis for computation models

The crucial part in developing a computation model is to obtain the basic material about customer opinions of the speech communication quality of the telephone service, in particular over the range of interest for the transmission parameters involved. There are several difficulties with this process.

A general problem is that opinions about the speech quality vary among people in the first place and also with time and with the context of the communication circumstances.

How the questions about quality are put will also influence the content of the answer. A classical example is that for listening levels. Experiments designed to establish "what is the most comfortable listening level" will yield a 3 dB lower result than if the task were to get answers to the question: "at what level is the speech quality highest?"

A straightforward method to obtain people's opinions is of course to *conduct customer surveys* in actual networks. However, during a telephone call people do not normally evaluate the speech transmission quality consciously but take a certain level for granted, in contrast to what is happening when listening to music on a hi-fi sound system. (Only if the telephone speech quality often falls below a certain threshold may the users react, and probably rather strongly, by complaints to the network operator.) Ideally, therefore a customer should be interviewed just after the completion of a call for which the transmission parameters were known more or less completely. Such investigations have been made, and are being made, in some instances even with wilfully introduced transmission impairments. The process is rather cumbersome and costly but should deliver the most reliable data.

The first step in the subjective evaluation of a transmission impairment is usually to employ a more economic method, in the form of *subjective tests under controlled conditions* in a laboratory. Such tests are made with a limited, specific set of transmission parameter variations and a limited number of test persons. (The set-up of such tests is described in the P-series Recommendations.) Most often, the result is presented as Mean Opinion Scores (MOS). However, great care is required in defining the test conditions and interpreting the results.

MOS values from a subjective test are often presented with statistical "confidence limits". However, it should be remembered that these limits only apply to that specific test occasion. One should not take an MOS value from a particular test as an absolute value as such but rather as a relative value. Thus, a carefully laid out subjective test should include reference conditions, i.e. the test persons are required also to judge impairments of a known nature. Using this reference MOS, the actual test MOS values are normalized so that results from different test laboratories can be compared (i.e. the test teams are "calibrated"). A Modulated Noise Reference Unit (MNRU) MOS is often used to produce the reference impairment. This has become the standard procedure when low-rate codecs are evaluated by subjective tests. (A source of uncertainty is that the "reference impairment" often sounds rather different from the "test impairment".)

As an example, Figure 1 shows four reference MNRU curves obtained by four different test laboratories in a common project to evaluate a certain low-rate LD-CELP codec (see Recommendation G.729). As can be seen, the spread in MOS values is up to almost 1.0 unit which indicates the need for normalizing the MOS values which the different laboratories measured for the actual codec under test.

All things considered, a computation model for predicting opinions must be tailored to a compromise of different subjective tests and customer surveys. The first step is to use results from laboratory subjective tests with a certain degree of "engineering judgement". A second step might be to check results from the model by comparisons with results obtained from customer surveys under real network conditions.

There are a number of opinion computation models described in the literature as is discussed in what follows. Considering the variability in subjective tests results and surveys, it should not come as a surprise if they were to predict slightly different results for the same network condition.

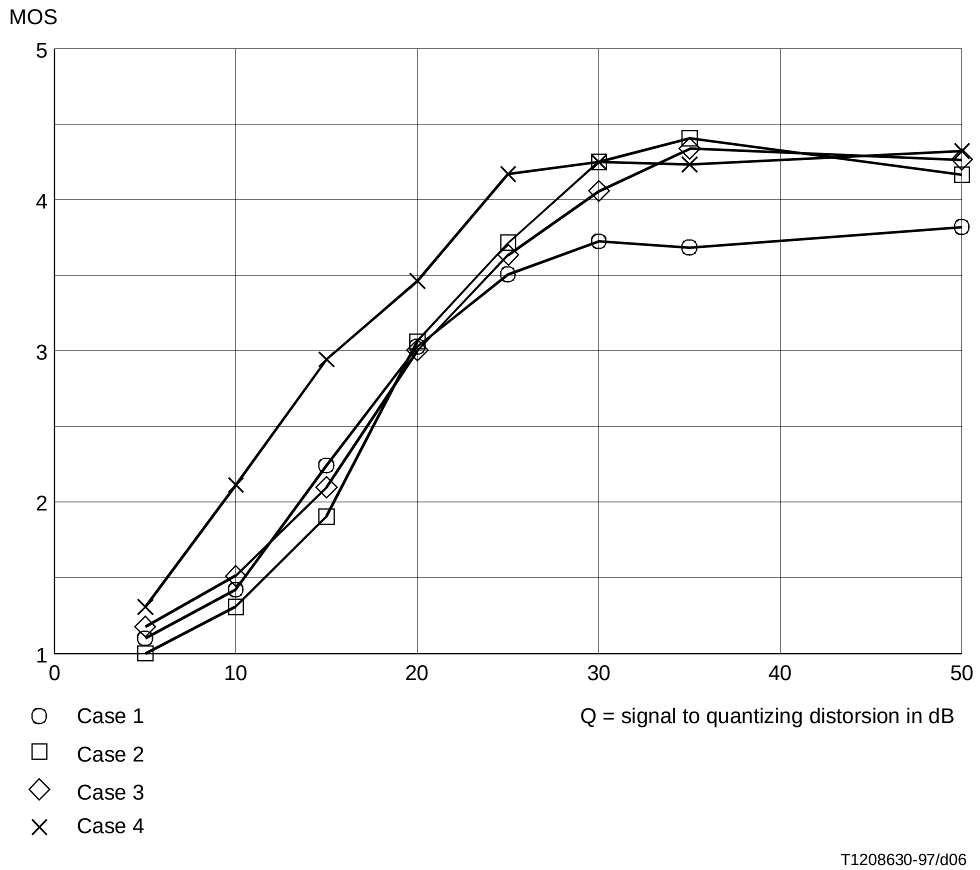


Figure 1 – Reference MOS curves, using MNRUs, from four different subjective test laboratories

3 Computation models as described in ITU-T publications

3.1 General

Supplement 3 (1993) to the P-series Recommendations contains descriptions of four different computation models for predicting transmission quality from objective measurements, namely:

– The “Transmission Rating” model as a contribution from USA and Canada.

– The “CATNAP83” model from British Telecom.

– The “Information Index” model as a contribution from France.

– The “OPINE” model from NTT.

Annex A/P.11 (*Blue Book*) describes the “Transmission Quality Index” model which is an amalgamated version of the four models in Supplement 4 to the P-series Recommendations, however with simplifications and a restricted number of parameters that are considered.

All these computation models were developed before the present trend of de-regulated telecom markets, i.e. when Administrations were able to control their networks in great technical details. Loss and noise were the major impairments and the influence of advanced low-rate codecs were not considered, at least not very accurately.

Appendix I/G.101 (1996) gives a description of the “E-model” which has been developed by a working group within ETSI, the European Telecommunications Standards Institute. It uses some algorithms and concepts from the models described in Supplement 3 to the P-series Recommendations complemented with results from recent subjective tests. Among the new features included are the subjective effects of low-rate codecs and the possibility to judge customers’ opinions of speech quality with regard to their expectation of the performance of the communication medium.

3.2 The “Transmission Rating” model

Parameters included are OR, circuit noise, room noise at the receive side, STMT, qdu (quantizing noise), bandwidth and attenuation distortion, listener echo and talker echo.

The computation results for a connection are presented as “R-ratings” which can be transformed into expected percentages of customers that find the connection “Good or Better” (GOB) or “Poor or Worse” (POW).

Of special interest are the following comments in the description of the model: “An important reason for the introduction of the R-scale was the recognition that subjective test results can be affected by various factors such as the subject group, the type of test, and the range of conditions which are included in the test. These factors have been found to cause changes in both the mean opinion score of a given condition and in the standard deviation. Thus, there are difficulties in trying to establish a unique relationship between a given transmission condition and subjective opinion in terms of mean opinion score or percent of ratings which are good or excellent. The introduction of a transmission rating scale tends to reduce this difficulty...”.

The subjective opinion data were gathered from laboratory tests as well as from rather elaborate surveys in the field.

Note that all the tests were conducted with Western Electric 500-type handsets or equivalent.

3.3 The “CATNAP83” model

The model deals with the subjective effects of circuit loss, attenuation/frequency distortion, circuit noise, quantizing noise, room noise and sidetone paths, for a reasonably wide range of values of these characteristics in any combination.

It is stated that the structure of this model makes the evaluation process reflect the cause-and-effect relationships which lead from the input (properties of the connection; acoustic environment; characteristics of the participants' hearing; speech sounds and language systems, etc.) to the output (participants' satisfaction or estimate of performance).

In practical use of the model (in the form of a computer program), connections need to be specified in terms of items and quantities such as noise levels, telephones of particular types, lengths of cables with stated resistance and capacitance per kilometre, and attenuators with stated loss. The program computes loudness ratings, speech levels and opinion scores. The latter are given both as the mean Listening Effort Score (Yle) and the mean Conversation Score (Yc).

3.4 The "Information Index" model

The theory for this model also considers the fundamental cause-and-effect features as CATNAP does. It takes into account transmission loss, circuit noise, room noise, attenuation/frequency distortion, sidetone and various distortions occurring in digital transmission.

The output from the model is the "Information Index" (I). In conjunction with the description of the model, examples are given, comparing calculated values of I with subjective MOS values.

3.5 The "OPINE" model

OPINE deals with transmission loss, circuit noise, attenuation/frequency distortion, quantizing distortion, talker echo and sidetone. It models the auditory-psychological process of evaluation by human beings of telephone performance based on these factors. (Thus, it is of the same general type as the CATNAP and Information Index models.)

Five psychological elements affecting telephone speech quality were chosen:

1) Speech distortion for attenuation/frequency distortion.

2) Effective loudness loss or excess in speech.

3) Noisiness during speech intervals and non-speech intervals.

4) Degradation caused by talker echo.

5) Degradation caused by sidetone.

Each psychological element is associated by a Performance Index (PI). The MOS for a connection is estimated from the sum of all PIs.

3.6 The "Transmission Quality Index" model

This, as stated in Supplement 3 to the P-series Recommendations, is a simple conversation opinion model for predicting the combined effects of OLR and psophometric noise. It also includes the effects of sidetone (STMR), room noise and attenuation distortion.

The model predicts MOS, percentages "Good and Excellent" and "Poor and Bad".

3.7 The "E-model"

In ETSI, the European Telecommunications Standards Institute, a working group has been working on a Technical Report on end-to-end speech transmission quality for telephone networks (designated ETR 250). Within that framework, a computational model has been developed for estimation of the network users' perception of the speech communication quality for a connection, using the network "conventional" transmission parameters as well as special "equipment impairment factors" for the codecs involved. The results have also been introduced in ITU-T in conjunction with revisions of Recommendations G.101 and G.113.

The ETSI model (the E-model) is to a large extent based on the “Transmission Rating” (TR) model described in 3.2, even if it uses some features from the other models in Supplement 3 to the P-series Recommendations. However, the structure is different from the TR model. The fundamental principle of the ETSI model is based on a concept established more than twenty years ago by J. Allnatt, and used for example in the NTT model OPINE: “Psychological factors on the psychological scale are additive”.

The ETSI model combines the effect of the various transmission parameters into a rating factor *R*, from which user reaction can be predicted, such as percentages finding the connection “Good or Better”, “Poor or Worse” or even so bad that they would terminate the call early, as well as what scores would be given in a MOS experiment. The rating factor *R* is composed of the terms:

$$R = Ro - Is - Id - Ie + A$$

where

Ro represents in principle the basic voice-signal-to-noise ratio;

Is, Id, Ie are so-called impairment factors;

Is represents impairments occurring simultaneously with the voice signal, like a too loud connection, loud sidetone and quantizing distortion from PCM;

Id represents delayed impairments, such as talker and listener echo as well as too long absolute delay;

Ie represents transmission impairments caused by special equipment such as certain low bit-rate codecs, Digital Circuit Multiplication Equipment, etc. (this factor is a new concept);

A is termed “Expectation Factor”. It represents an “advantage-of-access” that certain systems, in particular mobile telephone systems, have over conventional wire-bound communication systems. (The perception of “good quality” is intimately connected with how customers’ expectations are fulfilled, a subject area that is covered by such concepts as “usability”, “utility” and “attitudes”. The expectation factor is a new concept which has not previously been used in computation models.)

The E-model does not include the specific effects of attenuation distortion and telephone set frequency responses. The reason is that, in a de-regulated telecom market, the operators do not know these parameters very accurately so one has to be satisfied with the loudness rating planning values instead. Moreover, the users’ handling of the handset introduces anyhow a large variability in the earcap leakage frequency response.

In practice, the ETSI model gives about the same *R*-value as the TR model for the “conventional” impairments. Also, the TR model can be rearranged so that it has the same structure of additive terms for those impairments that are included in it. Results from application of the ETSI model on typical connections agree well with results from other models and from published subjective tests. In particular, impairments caused by low bit-rate codecs can be quite well predicted by the model, better than by the hitherto used methodology of “quantizing distortion units”. (In actual planning use, the most convenient form to use the E-model is to give limits to the sum of the impairment factors as is being done in Recommendation G.113.)

How to evaluate low bit-rate codecs, i.e. to transform subjectively derived MOS values into “equipment impairment factors”, is described in Annex E/G.113.

The additive structure of the E-model makes it simple to update if new evidence from subjective test with “old” parameters indicates it to be necessary, or new types of impairments need to be considered.

4 Future developments with regard to “cognitive and perceptual” models

This is an area where much research is going on and it is also a study point in ITU-T Study Group (SG) 12. The main reason is the need for developing reliable, objective methods for evaluating low-rate codecs (of which new types appear to be continuously introduced). At present, the performance of such codecs must be investigated by costly, subjective tests in order to obtain realistic and reliable results.