



INTERNATIONAL TELECOMMUNICATION UNION

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

P.56

(03/2011)

SERIES P: TERMINALS AND SUBJECTIVE AND
OBJECTIVE ASSESSMENT METHODS

Objective measuring apparatus

Objective measurement of active speech level

CAUTION !

PREPUBLISHED RECOMMENDATION

This prepublication is an unedited version of a recently approved Recommendation. It will be replaced by the published version after editing. Therefore, there will be differences between this prepublication and the published version.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, information and communication technologies (ICTs). The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure, e.g., interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU [had/had not] received notice of intellectual property, protected by patents, which may be required to implement this Recommendation. However, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database at <http://www.itu.int/ITU-T/ipr/>.

© ITU 2011

All rights reserved. No part of this publication may be reproduced, by any means whatsoever, without the prior written permission of ITU.

Recommendation ITU-T P.56

Objective measurement of active speech level

Summary

This Recommendation describes an easy to implement method to objectively determine the active speech level. It also acts as a reference against which other methods can be compared. The purpose of this Recommendation is not to exclude any other method but to ensure that results from different methods give the same result.

In the 2011 revision Annex B has been added providing the necessary information on how to use the active speech level meter for super-wideband (SWB) signals.

Recommendation ITU-T P.56

Objective measurement of active speech level

1 Scope

The ITU-T considers it important that there should be a standardized method of objectively measuring speech level, so that measurements made by different organizations may be directly comparable. Requirements of such a meter are that it should measure active speech level and should be independent of operator interpretation.

In this Recommendation, a meter is a complete unit that includes the input circuitry, filter (if necessary), processor and display. The processor includes the algorithm of the detection method. This meter can safely be used for laboratory experiments or can be used with care on operational circuits.

The method described herein maintains maximum comparability and continuity with past work, provided suitable monitoring is used, e.g. an operator performing the monitoring function. In particular, the new method yields data and conclusions compatible with those that have established the conventional value (22 microwatts) of speech power at the input to the 4-wire point of the international circuit according to Recommendation G.223. A method using operator monitoring can be found in Annex A.

This Recommendation describes a method that can be easily implemented. It also acts as a reference against which other methods can be compared. The purpose of this Recommendation is not to exclude any other method but to ensure that results from different methods give the same result.

Active speech level shall be measured and reported in decibels relative to a stated reference according to the methods described below, namely:

- *Method A* – Measuring a quantity called speech volume, used for the purpose of real-time control of speech level (see clause 4);
- *Method B* – Measuring a quantity called active speech level, used for other purposes (see clause 5).

Comparison of readings given by meters of methods A and B can be found in the *Handbook on Telephonometry*.

In the 2011 revision of the Recommendation Annex B has been added providing the necessary information on how to use the active speech level meter for super-wideband (SWB) signals.

NOTE – This meter cannot be used to determine peak levels but sufficient information exists giving the instantaneous peak/rms ratio, provided the signal has not been restricted or modified in any way, e.g. peak clipping.

2 Terminology

The recommended terminology is as follows:

<i>speech volume</i>	Until now used interchangeably with <i>speech level</i> , should in future be used exclusively to denote a value obtained by method A;
<i>active speech level</i>	Should be used exclusively to denote a value obtained by method B;

speech level Should be used as a general term to denote a value obtained by any method yielding a value expressed in decibels relative to a stated reference.

The definitions of these terms, and other related terms such as those for the meters themselves, should be adjusted accordingly.

3 General

3.1 Electrical, acoustic and other levels

This Recommendation deals primarily with electrical measurements yielding results expressed in terms of electrical units, generally decibels relative to an appropriate reference value such as one volt. However, if the calibration and linearity of the transmission system in which the measurement takes place are assured, it is possible to refer the result backwards or forwards from the measurement point to any other point in the system, where the signal may exist in some non-electrical form (e.g. acoustical). Power is proportional to squared voltage in the electrical domain, squared sound pressure in the acoustical domain, or the digital equivalent of either of these in the numerical domain, and the reference value must be of the appropriate kind (1 volt, 1 pascal, reference acoustic pressure equal to 20 micropascals, or any other stated unit, as the case may be).

3.2 Universal requirements

For speech-level measurements of all types, the information reported should include: the designation of the measuring system, the method used (A, B, or B-equivalent as explained in clause 4, or other specified method), the quantity observed, the units, and other relevant information such as the margin value (explained below) where applicable.

All the relevant conditions of measurement should also be stated, such as bandwidth, position of the measuring instrument in the communication circuit, and presence or absence of a terminating impedance. Apart from the stated band limitation intended to exclude spurious signals, no frequency weighting should be introduced in the measurement path (as distinct from the transmission path).

3.3 Averaging

Where an average of several readings is reported, the method of averaging should be stated. The *mean level* (mean speech volume or mean active speech level), formed by taking the mean of a number of decibel values, should be distinguished from the *mean power*, formed by converting a number of decibel values to units of power, taking the mean of these, and then optionally restoring the result to decibels.

Any correction that has been applied should be mentioned, together with the facts or assumptions on which any such correction is based. For example, in loading calculations, when the active levels or durations of the individually measured portions of speech differ widely, $0.115 \sigma^2$ is commonly added to the median or mean level in order to estimate the mean power, on the grounds that the distribution of mean active speech levels (dB values) is approximately Gaussian.

4 Method A – Immediate indication of speech volume for real-time applications

Measurement of speech volume for rapid real-time control or adjustment of level by a human observer should be accomplished in the traditional manner by means of one of the devices listed in Recommendation P.52.

The choice of meter and the method of interpreting the pointer deflexions should be appropriate to the application, as in Table 1.

Values obtained by method A should be reported as *speech volume*; the meter employed, the quantity observed, and the units in which the result is expressed, should be stated.

TABLE 1/P.56

Application	Meter	Quantity observed
Control of vocal level in live-speech loudness balances	ARAEN volume meter (SV3)	Level exceeded in 3 s
Avoidance of peak limiting	Peak programme meter	Highest reading
Maintenance of optimum level in making magnetic tape recording	VU meter	Average of peaks (excluding most extreme)

5 Method B – Active speech level for other applications than those mentioned in method A

5.1 Principle of measurement

Active speech level is measured by integrating a quantity proportional to instantaneous power over the aggregate of time during which the speech in question is present (called the active time), and then expressing the quotient, proportional to total energy divided by active time, in decibels relative to the appropriate reference.

The mean power of a speech signal when known to be present can be estimated with high precision from samples taken at a rate far below the Nyquist rate. However, the all-important question is what criterion should be used to determine when speech is present.

Ideally, the criterion should indicate the presence of speech for the same proportion of time as it appears to be present to a human listener, excluding noise that is not part of the speech (such as impulses, echoes, and steady noise during periods of silence), but including those brief periods of low or zero power that are not perceived as interruptions in the flow of speech. It is not essential that the detector should operate exactly in synchronism with the beginnings and ends of utterances as perceived: there may be a delay in both operating and releasing, provided that the total active time is measured correctly. For this reason, complex real-time voice-activity detectors depending on sampling at the Nyquist rate, such as those that have been successfully used in digital speech interpolation, are not necessarily the most suitable for this application. Their function is to indicate when a channel is available for transmission of information: this state does not always coincide with the absence of speech; on the one hand, it may occur during short intervals that ought to be considered part of the speech, and on the other hand, it may be delayed long after the end of an utterance (for reasons of convenience in the allocation of channels, for example).

This Recommendation describes the detection method that meets the requirements. The method involves applying a signal-dependent threshold which cannot be specified in advance, so that accurate results cannot be guaranteed while the measurement is actually in progress; despite that, by accumulating sufficient information during the process, it is possible to apply the correct threshold retrospectively, and hence to output a correct result almost as soon as the measurement finishes. Continuous adaptation of the threshold level in real time appears to yield similar results in simple cases, but further study is needed to find out how far this conclusion can be generalized.