International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

P.340
Amendment 1
(10/2014)

SERIES P: TERMINALS AND SUBJECTIVE AND OBJECTIVE ASSESSMENT METHODS

Voice terminal characteristics

Transmission characteristics and speech quality parameters of hands-free terminals

Amendment 1: New Annex B: Objective test methods for multi-talker scenarios

Recommendation ITU-T P.340 (2000) – Amendment 1

ITU-T

International
Telecommunication
Union

ITU-T P-SERIES RECOMMENDATIONS

TERMINALS AND SUBJECTIVE AND OBJECTIVE ASSESSMENT METHODS

Vocabulary and effects of transmission parameters on customer opinion of transmission quality	Series	P.10
Voice terminal characteristics	**Series**	**P.30** **P.300**
Reference systems	Series	P.40
Objective measuring apparatus	Series	P.50 P.500
Objective electro-acoustical measurements	Series	P.60
Measurements related to speech loudness	Series	P.70
Methods for objective and subjective assessment of speech quality	Series	P.80 P.800
Audiovisual quality in multimedia services	Series	P.900
Transmission performance and QoS aspects of IP end-points	Series	P.1000
Communications involving vehicles	Series	P.1100
Models and tools for quality assessment of streamed media	Series	P.1200
Telemeeting assessment	Series	P.1300
Statistical analysis, evaluation and reporting guidelines of quality measurements	Series	P.1400
Methods for objective and subjective assessment of quality of services other than voice services	Series	P.1500

For further details, please refer to the list of ITU-T Recommendations.

Recommendation ITU-T P.340

Transmission characteristics and speech quality parameters of hands-free terminals

Amendment 1

New Annex B: Objective test methods for multi-talker scenarios

Summary

The existing Recommendation ITU-T P.340 provides audio performance requirements and test methods for conference phones and hands-free terminals. These are limited to one-to-one meeting scenarios. Amendment 1 to Recommendation ITU-T P.340 introduces Annex B: Objective test methods for multi-talker scenarios.

History

Edition	Recommendation	Approval	Study Group	Unique ID*
1.0	ITU-T P.34	1980-11-21		11.1002/1000/8001
2.0	ITU-T P.34	1984-10-19		11.1002/1000/5905
3.0	ITU-T P.34	1988-11-25		11.1002/1000/1726
4.0	ITU-T P.34	1993-03-12	XII	11.1002/1000/1727
5.0	ITU-T P.340	1996-08-30	12	11.1002/1000/3632
6.0	ITU-T P.340	2000-05-18	12	11.1002/1000/5078
6.1	ITU-T P.340 (2000) Cor. 1	2004-03-31	12	11.1002/1000/7296
6.2	ITU-T P.340 (2000) Amd. 1	2014-10-29	12	11.1002/1000/12324

* To access the Recommendation, type the URL http://handle.itu.int/ in the address field of your web browser, followed by the Recommendation's unique ID. For example, http://handle.itu.int/11.1002/1000/11830-en.

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 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/.

Table of Contents

Recommendation ITU-T P.340

Transmission characteristics and speech quality parameters of hands-free terminals

Amendment 1

New Annex B: Objective test methods for multi-talker scenarios

1) Scope

This amendment adds new Annex B, Objective test methods for multi-talker scenarios, to Recommendation ITU-T P.340.

2) Modifications to ITU-T P.340

2.1) Clause 3

Add the following abbreviations to clause 3 in alphabetical order:

DUT Device Under Test

HATS Head And Torso Simulator

2.2) Annex B

Add Annex B with the following new text after Annex A:

Annex B

Objective test methods for multi-talker scenarios

(This annex forms an integral part of this Recommendation.)

B.1 Introduction

The existing standardized methods for measuring the performance of conference phones and hands-free terminals are dedicated mainly to one-to-one communications, i.e., the situation where only one person in a room is speaking, or when multiple people are speaking, turn-taking is so infrequent that the dynamics of turn-taking can be ignored.

However, the primary use case of conference phones and a frequent use case of hands-free terminals is to capture the voices of several people in a room. In a typical teleconferencing situation people may be speaking one at a time for a prolonged period, such as during a presentation, or they may be taking turns in rapid succession, such as during a discussion. When multiple talkers in a room are having interactive, natural conversations among themselves with a high level of interactivity, i.e., with sudden turn-taking or concurrent talking, there is a chance that remote participants cannot hear all talkers (sometimes without noticing) and may also hear distorted speech, due to the limited room capture-capability of conference phones.

This annex expands the scope to the communication situation where multiple talkers in a room interact frequently. Four performance criteria and their associated measurement techniques are defined. The first characterizes the adaptation time of the device under test (DUT) when it is presented with alternating (i.e., non-overlapping) talk bursts from different angular directions. The second quantifies how well the DUT preserves the level of each of two concurrent (i.e., completely overlapping) talk bursts. The third simulates more complicated turn-taking dynamics between two talkers including switching with partial overlap of talk bursts and abrupt talker switching without overlap. This criterion evaluates how well the level of a talker is maintained immediately after talker switching. The fourth criterion describes the concurrent talkers in the frequency domain to quantify how well the timbre of individual talkers is maintained.

The test methods described in this Annex use a set of artificial signals. The use of artificial signals such as composite source signal (CSS) may not be appropriate for some devices. The use of real speech signals is for further study.

B.2 Test setup and configuration

The test setup shall meet the requirements described in clause 5, [ITU-T P. 341] and [ITU-T P.581]. The head and torso simulator (HATS) should be calibrated per [ITU-T P.581], and should be positioned according to the test setup arrangements per clause 4 of [ITU-T P.341], e.g., 80 cm of horizontal and 30 cm of vertical distance from the edge of the DUT to the lip ring centre of the HATS for conference phones as shown in Figure 6 of [ITU-T P.341]. Test signal levels applied acoustically to the DUT should be calibrated following the procedures in clause 5. The receiving level of test signals, simulating the far-end communication point, should be set to −16 dBm0. If the DUT provides only multi-channel voice to its sending direction, the measurement signal at the electrical reference point should be down-mixed to mono.

The HATS should be arranged around the DUT with a certain angle setting throughout the tests. The angle terminology used in this Annex is depicted in Figure B.1. Using the reference zero-degree angle as the control panel of the conference phone, the angle value increases in a clockwise direction.

Because the performance of these tests may depend on the angular separation between the HATSs and the orientation of the HATSs towards the end point, the tests described in this Annex may need to be performed at multiple orientations to completely specify the DUT. It is the responsibility of the experimenter to select the HATS angles and orientations that are appropriate for the DUT. For DUTs that appear symmetric, it is recommended to test the combinations of angles A and B given in Table B.1.

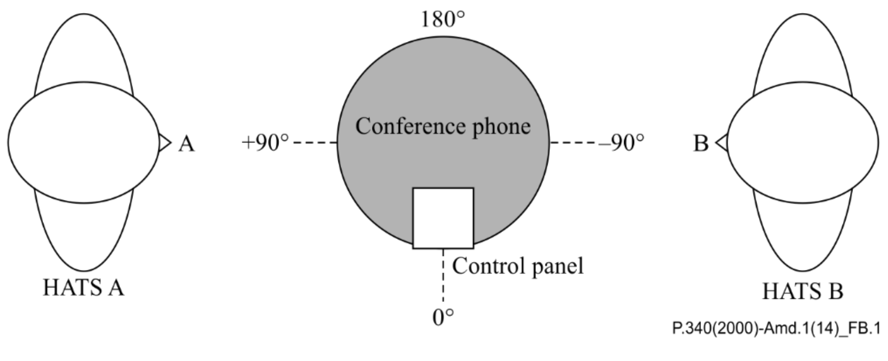

Figure B.1 – Diagram showing a conference phone and HATS angle configuration

HATS A and B are positioned at the angle of 90 degree and -90 degree, respectively, in Figure B.1. Talker angle at zero-degree refers to the direction of the control panel.

Table B.1 – Recommended HATS angle configurations

No	Angle A [degree]	Angle B [degree]		No	Angle A [degree]	Angle B [degree]
1	0	45		13	90	45
2	0	90		14	180	−135
3	0	135		15	180	−90
4	0	180		16	180	−45
5	0	−135		17	180	45
6	0	−90		18	180	135
7	0	−45		19	−90	−45
8	90	135		20	−90	45
9	90	180		21	−90	135
10	90	−135		22	−90	−135
11	90	−90		23	−90	−45
12	90	−45				

B.3 Test 1: Adaptation time in talker alternation

Purpose

The purpose of this test is to measure the adaptation performance of the DUT to two alternating (i.e., non-overlapping) talk bursts, applied from different angles.

Method

A full-band CSS and its double talk version are used in this test [ITU-T P.501]. The sampling rate should be 44.1 kHz or 48 kHz. The level of the test signal should be −4.7 dBPa at the MRP. As shown in Figure B.2, the input signal is constructed such that a pair of two CSSs is applied to the HATS positioned at angle A (upper channel) and a pair of double talk CSSs is applied to another HATS located at angle B (lower channel) in an alternating manner. The signal shown in Figure B.2 (1.6 s duration) is repeated eight times to create the test signal. This simulates two alternating talkers sitting at different angles around the DUT.

The measurement should be done at the electrical reference point. For each period of the test signal (Figure B.2), L_A is calculated as the average of two RMS levels, each of which is measured over the active portion of the pair of CSSs (excluding the pause) corresponding to the angle A. Given a series of L_As over multiple periods of the measurement signal, i.e., $L_A(n)$, $n = 0, 1, \ldots, 7$, the target adaptation level, TAL_A, is defined as $L_A(n)$ with the smallest n, if both $|L_A(n) - L_A(n\text{-}1)|$ and $|L_A(n) - L_A(n\text{-}2)|$ are less than 1 dB. The adaptation period is defined as the smallest n such that $| L_A(n) - TAL_A | < 1$. Similarly, L_B, the target adaptation level and adaptation period of angle B, can be defined in the same manner.

These measurements should be repeated with different levels of test signal at angle B (−6 dB, −3 dB, 3 dB and 6 dB relative to the original level at angle B) and different combinations of angles A and B.

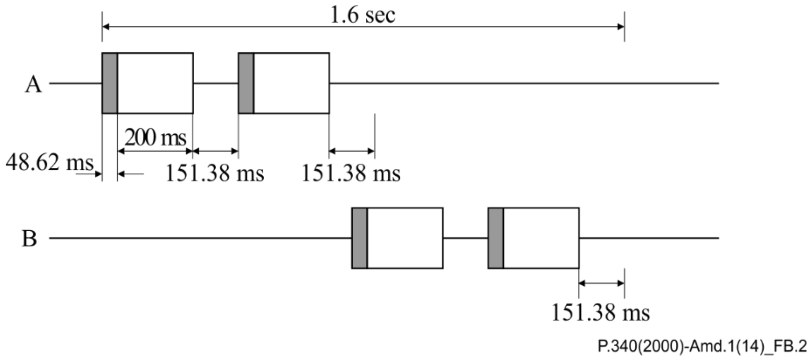

Figure B.2 – One period of input signal for Test 1

The upper channel and lower channel are applied at angle A and B, respectively. The signal simulates two alternating talkers sitting at angle A and angle B.

Requirement

To ensure a stable state for subsequent measures, the DUT shall asymptote to a stable state and shall not deviate from that target adaptation level by more than 1 dB. This requirement must be satisfied for both angles A and B.

Guidance

A faster adaptation time is preferable to a slower adaptation time. The adaptation period should be equal to or less than 1 period (i.e., the DUT should be adapted in 1.6 s for the test signal characterized in Figure B.2).

Experimenters are recommended to report the actual measurement results of $L_A(n)$ and $L_B(n)$ as well as the quantities described as the requirements.

B.4 Test 2: Level of completely overlapping (concurrent) talk bursts

Purpose

The purpose of this test is to measure the capture performance of the DUT, when two different talk bursts, applied from different angles, are completely overlapped in time. The level of concurrent talk bursts should reflect the levels of both single talk bursts.

Method

After the DUT is adapted according to Test 1, the concurrent signal shown in Figure B.3 is applied which consists of a pair of CSSs at angle A and a pair of double talk CSSs at angle B. The average of two RMS levels, measured at the electrical reference point over the active portion of the pair of CSSs (excluding the pause), is calculated for each interval: L_A for single talk A, L_B for single talk B, and L_C for concurrent talk. The level of concurrent talk bursts, L_c, is measured in relation to the level of the preceding single talk bursts A and B.

These measurements should be repeated with different levels of test signals at angle B (−6 dB, −3 dB, 3 dB and 6 dB relative to the original level at angle B) and different combinations of angles A and B.

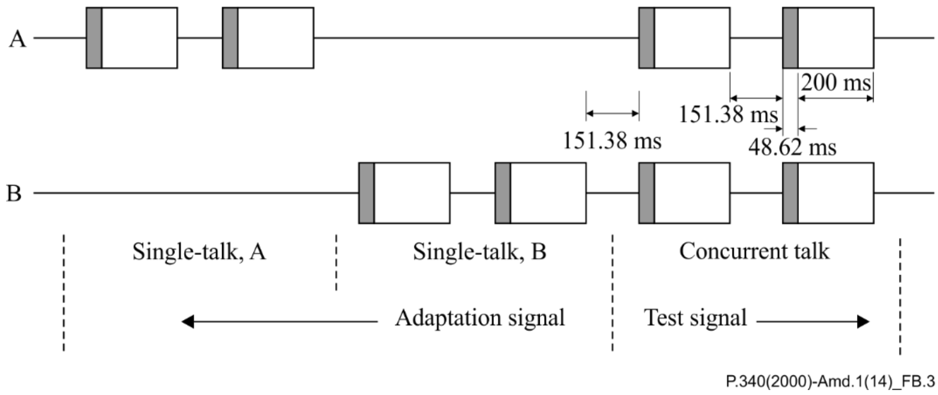

Figure B.3 – Input signal of Test 2

The upper and lower channel are applied at angle A and B, respectively. The last period of the adaptation signal is shown for the angle A and B, followed by the concurrent talk bursts. The level of concurrent talk bursts is measured in relation to the level of preceding single talk bursts A and B.

Guidance

In a transparent system, such as an air path, L_C equals L_D, which is the energetic sum of L_A and L_B:

$$L_D = 10\log\left(10\hat{}(L_A/10) + 10\hat{}(L_B/10)\right) \quad \text{(B.1)}$$

When L_C does not equal L_D, a smaller absolute difference is preferable to a large one. Differences less than 1 dB are considered perceptually irrelevant. L_C and L_D should be reported as a minimum.

B.5 Test 3: Dynamic turn-taking - switching characteristics

Purpose

The purpose of this test is to measure the performance of the DUT to the dynamic turn-taking between two talkers from different angles. The level measured immediately after talker switching should not demonstrate major distortions. Two different switching scenarios are covered in this test: 1) partially-concurrent talk bursts and 2) abrupt talker switching without overlap.

Method

After the DUT is adapted according to Test 1, a sequence of CSSs at angle A and a sequence of double talk CSSs at angle B are applied as shown in Figure B.4. In the test signal, the interval A_X represents an active talker switching from angle B to angle A with partial overlap of CSS burst. The interval B_X corresponds to an abrupt switching of a talker from angle A to angle B without any overlap in time. The level over time, measured at the electrical reference point, is calculated using an exponential weighting filter with a time constant of 5 ms [ITU-T G.168]. The time series of level values in the interval A_X ($L_{A,X}$) is compared with the time series of level values in the interval A_{REF}, which is the corresponding part of A_X when no talker switching happens. In the same manner, the level values in the interval B_X are compared with those in the interval B_{REF}.

These measurements should be repeated with different levels of test signals at angle B (−6 dB, −3 dB, 3 dB and 6 dB relative to the original level at angle B) and different combinations of angles A and B.

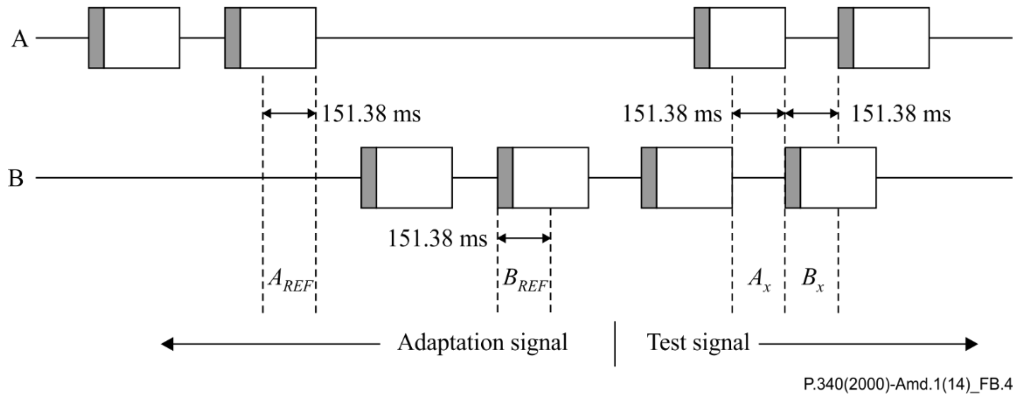

Figure B.4 – Input signal of Test 3

The upper channel and lower channel are applied at angle A and B, respectively. The interval A_X represents an active talker switching from angle B to angle A with partial overlap of burst in time, and the interval B_X represents an abrupt switching of a talker from angle A to angle B without any overlap of bursts in time.

Guidance

Performance is evaluated on both long and short time scales.

- average level – the loss of signal level after talker switching in average sense:

  In an ideal (transparent) system the difference between the average level in A_X and A_{REF} (and B_X and B_{REF}) should be 0 dB. A device with a smaller deviation of the absolute difference is preferable to one with a larger absolute difference. Devices with absolute deviations less than 3 dB deliver excellent switching performance.

- minimum level – the loss of signal level after talker switching in short time sense:

  In an ideal (transparent) system the difference between the minimum level in A_X and A_{REF} (and B_X and B_{REF}) should be 0 dB. A device with a smaller difference between these minimum levels is preferable. Devices with differences between the minimum levels less than 6 dB deliver excellent switching performance.

B.6 Test 4: Concurrent talk test with voice-like AM-FM signals

Purpose

The purpose of this test is to measure the capture performance of the DUT in maintaining frequency contents of two concurrent input signals applied from different angles.

Method

To enable the separation of two input signals, after processing by the DUT, two voice-like signals modulated in amplitude and frequency for wideband applications (S_A and S_B) are used in this test, see clause 7.2.4 of [ITU-T P.501]. These input signals are designed such that the frequency contents of the two signals do not overlap, as seen in Figure B.5. S_A, has 29 frequency components starting from 125 Hz and S_B has 28 frequency components starting from 150 Hz.

After the DUT is adapted according to Test 1, the first voice-like signal, S_A, is applied to the HATS at angle A, followed by a pause (> 0.5 s) and the second voice-like signal, S_B, is applied to the HATS at angle B. Then both S_A and S_B are applied in a concurrent manner. This results in three time intervals as shown in Figure B.6: 'Single talk A', 'Single talk B' and 'Concurrent talk'.

The measured signal during the 'Single talk A' interval is processed by filter A (blue dotted frequency response in Figure B.5) to select the frequency contents of S_A, and the frequency response gain,

$G_{A,single}(k)$, is calculated for each k-th frequency band. The measured signal during the 'Single talk B' interval is processed by filter B (red dotted frequency response in Figure B.5) to calculate the frequency response gain, $G_{B,single}(k)$ for each k-th frequency band. These two gains represent the frequency response of the DUT for single talk (no overlap), and are used as references when evaluating the response of the concurrent talkers case.

From the measured signal during the 'Concurrent talk' interval, two different gains are estimated. The first is the concurrent gain for S_A, $G_{A,con}(k)$, obtained through filter A for each frequency band. Similarly, the concurrent gain for S_B, $G_{B,con}(k)$, is obtained using filter B.

The measurements should be repeated with different levels of test signals at angle B (−6 dB, −3 dB, 3 dB and 6 dB relative to the original level at B) and different combinations of angles A and B.

Guidance

In an ideal, transparent system the signal components of signal A do not affect the signal components of signal B and vice versa. Therefore, in such a system the concurrent gain of all the corresponding frequency bands should not deviate more than 0 dB from the single talk gain.

$$|GD_A(k)| = |G_{A,con}(k) - G_{A,single}(k)| = 0 \text{ dB for all } k = 1, 2, \ldots, 29 \quad \text{(B.2)}$$

where:

k is the frequency band index for signal A

$$|GD_B(k)| = |G_{B,con}(k) - G_{B,single}(k)| = 0 \text{ dB for all } k = 1, 2, \ldots, 28 \quad \text{(B.3)}$$

where:

k is the frequency band index for signal B

When deviations do occur, DUTs with smaller absolute values of these differences are considered to have better timbre characteristics than DUTs with larger differences. DUTs with absolute differences less than 3 dB in all bands are considered to have excellent performance during concurrent talk.

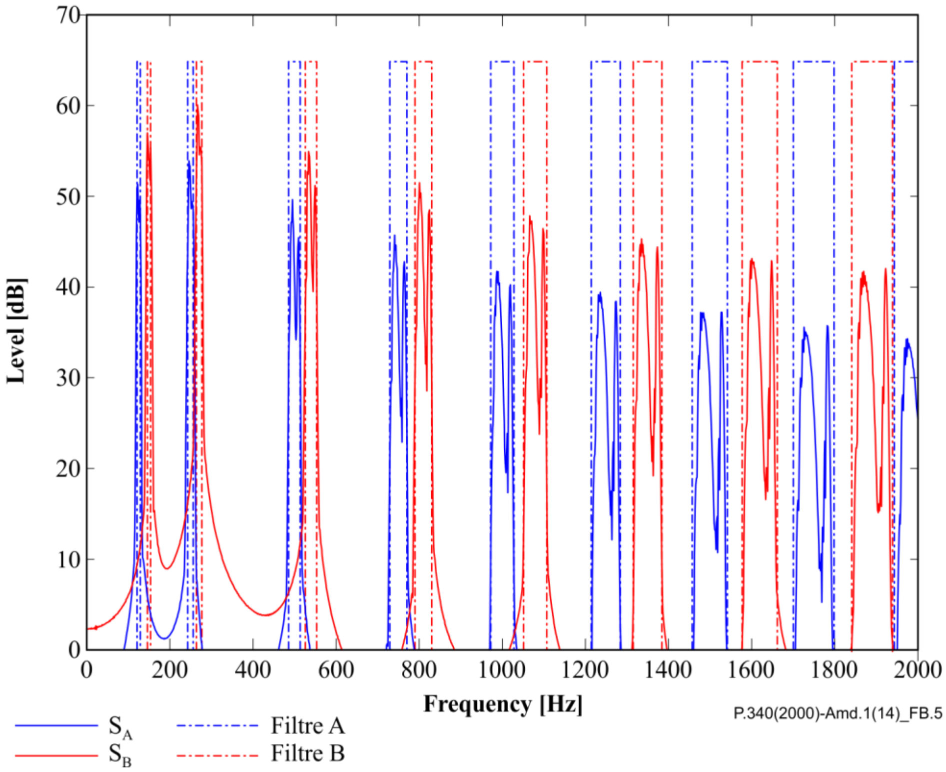

Figure B.5 – Power spectral density of input signals for Test 4 (shown only up to 2 kHz, although the signal contents span to 7 kHz range)

In Figure B.5 the spectrum in blue shows frequency components of the signal applied at angle A (S_A), and the red frequency components are for the signal applied at angle B (S_B). The blue dotted line represents the filter response used for the analysis of S_A, and the red dotted line is for the analysis of S_B.

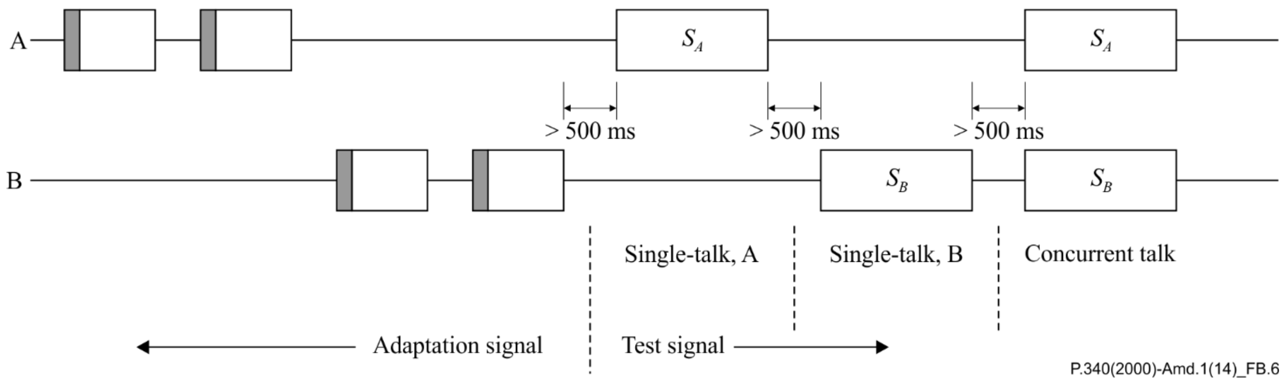

Figure B.6 – Input signal diagram of Test 4

The upper and lower channel are applied at angle A and angle B, respectively. The last period of the adaptation signal is shown for angle A and angle B, followed by two single talk intervals and a concurrent talk interval.

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
Series D	General tariff principles
Series E	Overall network operation, telephone service, service operation and human factors
Series F	Non-telephone telecommunication services
Series G	Transmission systems and media, digital systems and networks
Series H	Audiovisual and multimedia systems
Series I	Integrated services digital network
Series J	Cable networks and transmission of television, sound programme and other multimedia signals
Series K	Protection against interference
Series L	Construction, installation and protection of cables and other elements of outside plant
Series M	Telecommunication management, including TMN and network maintenance
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	**Terminals and subjective and objective assessment methods**
Series Q	Switching and signalling
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminals for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	Data networks, open system communications and security
Series Y	Global information infrastructure, Internet protocol aspects and next-generation networks
Series Z	Languages and general software aspects for telecommunication systems

Printed in Switzerland
Geneva, 2015