

# **AQM in TEMS Products (Including PESQ)**

Technical Paper

© Ascom 2009. All rights reserved.

TEMS is a trademark of Ascom. All other trademarks are the property of their respective holders.

No part of this document may be reproduced in any form without the written permission of the copyright holder.

The contents of this document are subject to revision without notice due to continued progress in methodology, design and manufacturing. Ascom shall have no liability for any error or damage of any kind resulting from the use of this document.

# Contents

<b>1. General .....</b>	<b>1</b>
1.1. Summary of Especially Advanced or Unique AQM Features ..	3
1.2. Organization of This Document.....	3
1.3. Definitions .....	4
1.4. AQM Parameters in TEMS Products .....	4
<b>2. The PESQ Algorithm.....</b>	<b>6</b>
2.1. Overview.....	6
2.2. PESQ in More Detail.....	6
2.3. Advantages of the PESQ Algorithm .....	8
2.4. PESQ Output: MOS-LQO .....	8
2.5. Speech References Used in PESQ .....	9
2.5.1. Impact of Speech Reference Content and Speech Coders on the PESQ Score.....	10
2.6. Intermediate Results from PESQ: Frequent AQM.....	11
2.6.1. How Close Is Frequent AQM to PESQ? .....	12
<b>3. PESQ in TEMS Automatic.....</b>	<b>13</b>
3.1. MTUs .....	14
3.1.1. Audio Connection.....	14
3.1.2. Monitoring the Audio .....	14
3.2. Call Generator .....	15
3.2.1. Audio Connection.....	15
3.2.2. Monitoring the Audio .....	15
3.3. Mobile Receiving Unit (MRU).....	15
3.4. Operator Console.....	15
3.5. PESQ Post-processing in TEMS Presentation.....	16
<b>4. PESQ in TEMS Investigation .....</b>	<b>17</b>
4.1. TEMS Investigation AQM Modules .....	18
4.1.1. Audio Connection.....	18
4.1.2. Monitoring the Audio .....	18
4.2. Call Generator .....	19
4.2.1. Audio Connection.....	19
4.2.2. Monitoring the Audio .....	19
4.3. Mobile Receiving Unit (MRU).....	19
4.4. PESQ Post-processing in TEMS Investigation .....	19
<b>5. Mobile-to-mobile PESQ.....</b>	<b>20</b>
5.1. Mobile Receiving Unit .....	20
5.2. AQM Modules .....	21
5.2.1. TEMS Automatic.....	21
5.2.2. TEMS Investigation.....	21

- 5.3. PESQ Post-processing in TEMS Presentation.....21
- 5.4. How to Use Mobile-to-mobile PESQ Measurement .....22
  - 5.4.1. Impact of Double Speech Coding in Mobile-to-mobile Calls on the PESQ Score.....22
- 6. PESQ Measurement Procedure.....24**
  - 6.1. Call Generator/AQM Module Measurement Procedure .....24
    - 6.1.1. Realtime Synchronization .....25
    - 6.1.2. Details of the Synchronization Algorithm .....26
  - 6.2. Mobile-to-mobile Measurement Procedure .....27
    - 6.2.1. Realtime Synchronization .....28
    - 6.2.2. Details of the Synchronization Algorithm .....28
    - 6.2.3. Mobile-to-mobile Specific Synchronization .....29
  - 6.3. Situations Where No PESQ Score Is Produced.....29
- 7. Verification of PESQ in TEMS Tools .....30**
  - 7.1. Step 1: Accuracy of P.862.1 .....31
  - 7.2. Step 2: Conformance Tests .....31
  - 7.3. Step 3: Live Tests .....32
  - 7.4. Verification of Frequent AQM .....33
- 8. Echo Measurements .....35**
  - 8.1. Overview of Echo Measurement.....35
  - 8.2. Some Details of the Echo Measurement Algorithm.....35
    - 8.2.1. Calculation of Echo Delay.....35
    - 8.2.2. Calculation of Echo Attenuation.....36
- 9. Silent Call Detection .....37**
  - 9.1. Some Details of Silent Call Detection .....37
    - 9.1.1. Minimum Call Duration .....37
    - 9.1.2. Silent Call Detection on Uplink .....37
    - 9.1.3. Silent Call Detection on Downlink.....38
    - 9.1.4. Call Not Synchronized .....38
- 10. Volume Measurements .....39**
  - 10.1. Factors Affecting Volume Measurements.....39
    - 10.1.1. Dependency on Sites.....39
    - 10.1.2. Dependency on Mobiles .....39
    - 10.1.3. Dependency on Speech Recordings .....39
    - 10.1.4. Radio Disturbances.....39
- 11. Speech Path Delay Measurements .....40**
  - 11.1. Factors Affecting Delay .....40
  - 11.2. Details on the Speech Path Delay Measurement.....40
  - 11.3. How to Use the Speech Path Delay Measurement .....41





# 1. General

The purpose of this document is to describe how audio quality measurements are conducted in TEMS Automatic and TEMS Investigation.

Audio quality measurement (AQM) in TEMS products are based on the unique combination of Ericsson<sup>1</sup> research competence, handset technology, and state-of-the-art testing facilities. The latter include a complete base station–MSC system, thus guaranteeing the correctness of audio quality measurements. Ericsson research efforts in the audio quality measurement area have centered on SQI on one hand and PESQ (ITU P.862.1) on the other.

In the table below is laid out the evolution of audio quality measurement in TEMS products.

Year	AQM Feature	Description
1997	SQI	Speech quality estimated from radio parameters. A non-intrusive measurement – not dependent on test signals. Research into the SQI algorithm started as early as 1995.
1998	PSQM	Speech-based speech quality measurement on both uplink and downlink using ITU P.861.
2000	Echo, volume, silent calls	Testing of echo canceller efficiency and network volume.
2003	Transform to MOS-LQO	Transform from PESQ MOS to MOS-LQO, designed to give scores comparable to listening tests. Collaboration with partners in ITU resulting in the ITU P.862.1 standard.
2004	PESQ	Speech-based speech quality measurement on both uplink and downlink using ITU P.862.1: implementation in TEMS Automatic.
2007	PESQ	Speech-based speech quality measurement on both uplink and downlink using ITU P.862.1: implementation in TEMS Investigation.
2008	M2M PESQ	Mobile-to-mobile PESQ. Speech-based speech quality measurement on downlink using ITU P.862.1: implementation in both TEMS Automatic and TEMS Investigation.

<sup>1</sup> The TEMS business was part of the Ericsson group until 2009, when it was acquired by Ascom. The role played by Ericsson in developing TEMS audio quality measurement is discussed in several places in this document.



## 1.1. Summary of Especially Advanced or Unique AQM Features

Below is a list of especially noteworthy features of AQM in TEMS Automatic and TEMS Investigation, some of which are unique and cannot be found in competing products.

- To obtain the best possible agreement with listening tests, the ITU standardized MOS-LQO scale is used: see section 2.4.
- To achieve better geographical resolution of speech quality problems, intermediate results from the PESQ algorithm are presented. This feature is called Frequent AQM; see section 2.6.
- Each mobile measurement probe calling a Call Generator allows the user to listen live to the downlink speech and simultaneously watch how the PESQ algorithm grades the speech. See sections 3.1.2 and 4.1.2.
- The speech sent on the uplink to the Call Generator is stored on file whenever a user-defined PESQ threshold is exceeded. These recordings can be further analyzed with a number of publicly available tools to give extensive insight into specific speech quality problems. This feature is described in sections 3.2.2 and 4.2.2.
- With TEMS tools one can determine whether a speech quality problem is located in the fixed part of the network or in the mobile part. This is done by comparing PESQ results with SQI (a unique TEMS feature for measuring speech quality based on cellular network parameters).
- The TEMS implementation of PESQ has been verified in a controlled base station–MSC environment. These advanced testing facilities are unique. The testing is described in chapter 7.

## 1.2. Organization of This Document

Audio quality measurements comprise PESQ, Frequent AQM, echo measurements, volume measurements, and silent call detection. Chapters 2–7 are concerned with PESQ and Frequent AQM:

- Chapter 2 describes the workings of the PESQ algorithm.
- Chapters 3 and 4 describe the implementation of PESQ in TEMS Automatic and TEMS Investigation respectively.
- Chapter 5 describes mobile-to-mobile PESQ.
- Chapter 6 goes into details of the PESQ measurement procedure.
- Chapter 7 relates how the PESQ algorithm has been verified.

The remaining chapters (8–11) deal with the other audio quality measurements.

### 1.3. Definitions

Term	Description
AQM	Audio quality measurements. See section 1.4 for an overview of audio quality measurements in TEMS products.
AQM module	Equipment used to obtain audio quality measurements. In TEMS Automatic, the AQM module is integrated into the MTU. In TEMS Investigation, AQM modules are either standalone or mounted in TEMS equipment cases.
PESQ	Perceptual Evaluation of Speech Quality: an objective method for estimating the subjective quality of cellular networks. It is described in the ITU recommendation P.862.

### 1.4. AQM Parameters in TEMS Products

The table below shows the AQM parameters that are measured by TEMS products.

TA = TEMS Automatic  
 TI = TEMS Investigation

Parameter	Call Generator		MRU
	Analog Calling Card	Digital Calling Card	
PESQ Score DL	TA + TI	TA + TI	TA + TI
PESQ Score UL	TA + TI	TA + TI	-
Frequent AQM DL	TA + TI	TA + TI	TA + TI
Frequent AQM UL	TA + TI	TA + TI	-
Volume DL	TA + TI	TA + TI	TA + TI
Echo Attenuation DL	TA + TI	-	-
Echo Correlation DL	TI	-	-
Echo Delay DL	TA + TI	-	-
Echo Power DL	TI	-	-
PESQ Sentence DL	TA + TI	TA + TI	TA + TI

Parameter	Call Generator		MRU
	Analog Calling Card	Digital Calling Card	
PESQ Sentence UL	TA + TI	TA + TI	–
Frequent AQM Sentence DL	TA + TI	TA + TI	TA + TI
Frequent AQM Sentence UL	TA + TI	TA + TI	–
Silent Call DL	TA + TI	TA + TI	TA + TI
Silent Call UL	TA	TA	–
Speech Path Delay <sup>2</sup>	TA + TI	TA + TI	–

<sup>2</sup> Recorded by the Call Generator as part of uplink AQM data, but actually concerns both links.

## 2. The PESQ Algorithm

The purpose of this section is to give a general understanding of how the PESQ algorithm works.

### 2.1. Overview

PESQ, Perceptual Evaluation of Speech Quality, is an objective method for estimating subjective speech quality in a cellular network. To put it differently, PESQ is a method to calculate how the cellular network subscriber experiences the quality of speech reproduction in the network. PESQ is described in the ITU recommendation P.862.

The aim of PESQ is to mimic the sound perception of humans. It assesses the quality of a distorted speech signal (i.e. one that has been coded and transmitted over the network) by comparing it with the original, undistorted signal:

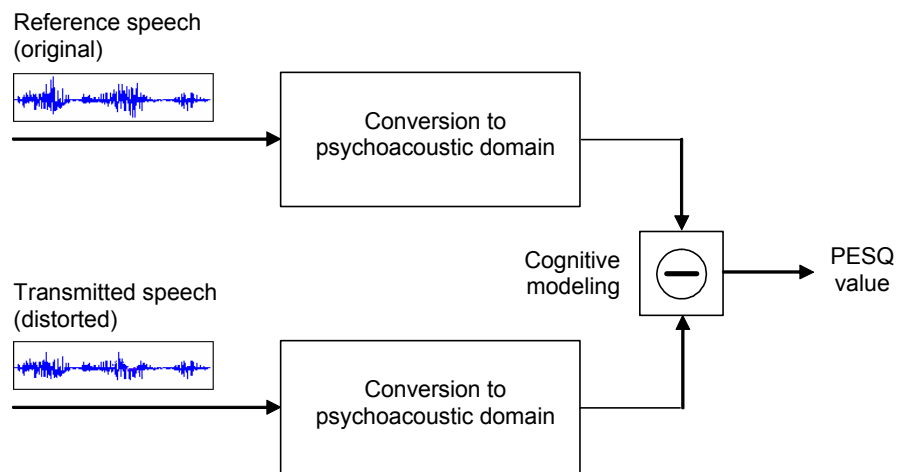


Figure 1 Overview of PESQ algorithm.

In PESQ, the original and the distorted speech are mapped onto psychophysical representations of how humans experience speech. The quality of the distorted speech is judged on the basis of differences in the psychophysical representations.

The PESQ algorithm produces a value ranging from 4.5 down to 1. A PESQ value of 4.5 means that the measured speech has no distortion: it is exactly the same as the original. A value of 1 indicates the severest degradation.

### 2.2. PESQ in More Detail

The PESQ algorithm consists of two parts: 1) conversion to the psychoacoustic domain and 2) cognitive modeling. The most important steps in each part are depicted in Figure 2 below.

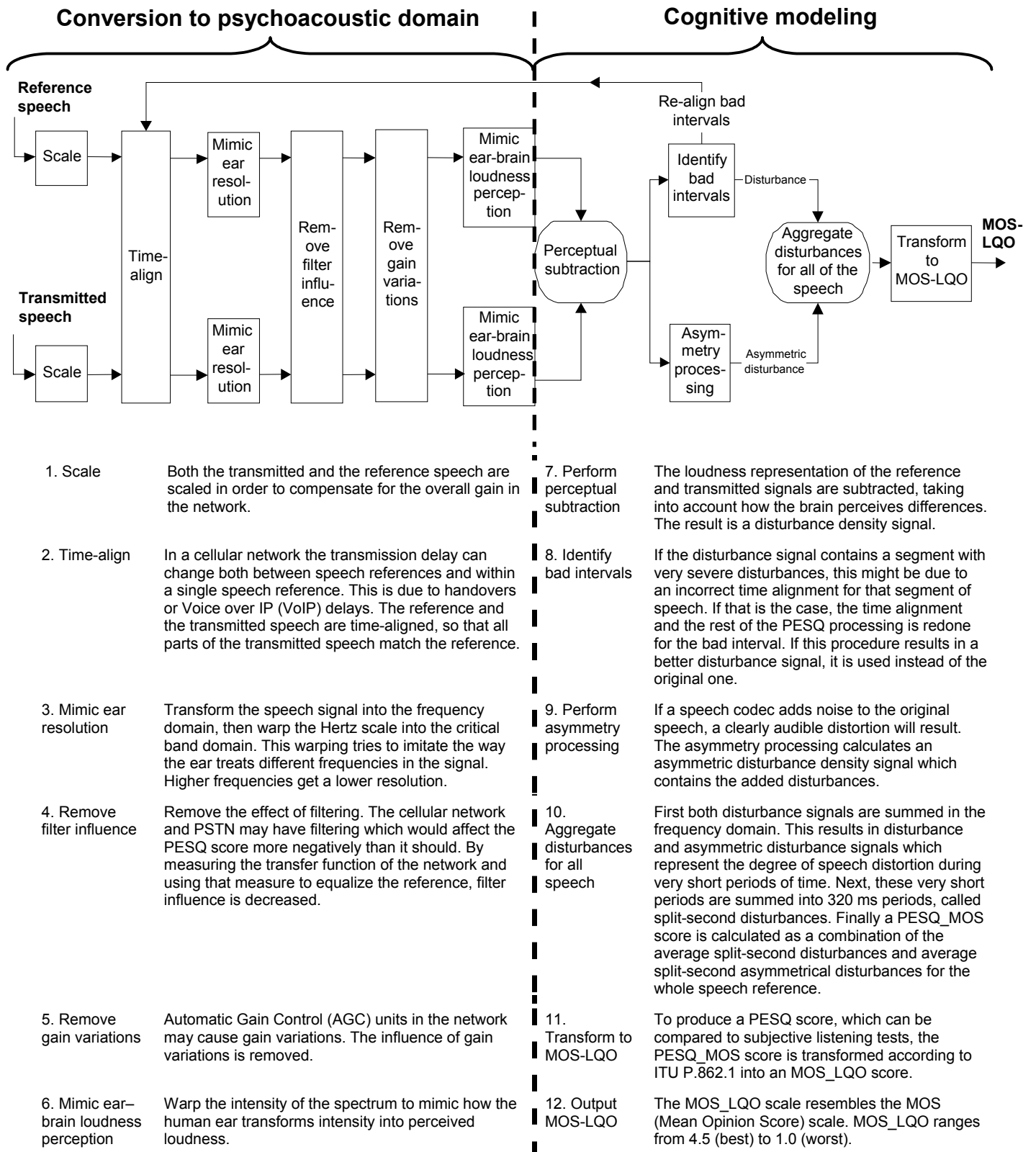


Figure 2 Detailed flowchart of PESQ algorithm.

### 2.3. Advantages of the PESQ Algorithm

When PESQ was introduced it was a major improvement over earlier algorithms (like PSQM). The most important improvements in the PESQ algorithm are that PESQ has a time alignment procedure which handles VoIP and handovers, and that PESQ removes effects of filtering. The cognitive modeling is also better in PESQ, resulting in slightly better correlations to subjective listening tests. The table below summarizes the advantages of PESQ.

Area	Comment
VoIP	PESQ handles some delay variations (jitter).
AMR, lower rates	PESQ handles the filtering applied by the lower rate AMR codecs.
Handovers	PESQ handles the short delays caused by handovers.
Filtering	PESQ handles filtering which could be performed by the PSTN or a company exchange.
Overall	PESQ won the ITU competition showing the best correlation to subjective listening tests.

### 2.4. PESQ Output: MOS-LQO

The output from the PESQ algorithm is given in terms of MOS-LQO. The use of this unit merits some words of explanation.

In subjective evaluations of speech coders, the Mean Opinion Score (MOS) scale is frequently used. The MOS scale is described in ITU recommendation P.800. In a MOS test the listener grades each speech sample (normally five to eight seconds in length) by assigning it to one of the following categories:

MOS	Meaning
5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

These ratings are defined solely by the words in the table, so that listeners must decide for themselves what constitutes an “excellent” (or “good”, etc.) speech sample. The loose definition of the MOS scale makes it sensitive to changes in the listening procedure. The MOS values obtained in a test depend on the background of the listeners, the quality range of the material, the loudness level, the listening equipment, and more. This means that MOS scores from different tests of the same telephone transmission system can vary slightly. For example, two different MOS tests with the same AMR speech coder will not give exactly the same result.

Nevertheless, since the MOS scale is the industry standard, it is desirable for measurement tools such as TEMS Automatic and TEMS Investigation to produce scores resembling those that would be obtained in a MOS listening test. This is accomplished by the transform into MOS-LQO, described in the ITU P.862.1 standard. MOS-LQO stands for Mean Opinion Score – Listening Quality Objective. The MOS-LQO scale goes from 4.5 (best) to 1.0 (worst).

## 2.5. Speech References Used in PESQ

The speech references used for PESQ in TEMS products have been chosen for their phonetic characteristics. A total of four references are used, featuring two male and two female speakers. Each reference consists of two sentences with a short period of silence in between them. The total length of each reference is 5.5 seconds, corresponding to 44,000 samples. Each sample is stored as a 16-bit value. This structure of the references is in accordance with the ITU standard P.830, which stipulates how to perform subjective MOS tests. The waveform of the reference no. 1 is shown in the graph below.



Figure 3 Waveform of reference no. 1. Each reference consists of two sentences.

### 2.5.1. Impact of Speech Reference Content and Speech Coders on the PESQ Score

Speech coders are sensitive to speaker pitch and to the formant structure of the speech (both of which differ, for example, between males, females and children). Coders are also sensitive to the content of a speech reference. The most important factor is the amount of silence in the reference, since the speech coder will introduce no degradation in silent parts. This means that speech coders will distort different speech references to different degrees. Silent parts are also not affected by radio disturbances as much as speech. In summary, therefore, the more silence the reference contains, the higher it will score. These factors together will cause different speech references to receive different PESQ values, even when the same speech coder is used, and regardless of the radio conditions.

The choice of speech coder (for example EFR, FR, AMR, or EVRC) has an even greater impact on the PESQ scores than the choice of speech reference.

The table below shows the difference in the PESQ score for the references used in TEMS Automatic. The table also illustrates the differences between two speech coders: EFR (used in GSM) and EVRC (used in CDMA). These results isolate the degradations caused by the speech coders – the radio channel is perfect, and no PSTN is involved. As we can see, for GSM using EFR, a variation of  $\pm 0.05$  can be expected between the references. For CDMA using EVRC, the variation is up to  $\pm 0.2$ . The average difference between these speech coders is 0.4.

Speech Coder	Reference No.	PESQ (MOS-LQO)
EFR	0	4.24
EFR	1	4.18
EFR	2	4.10
EFR	3	4.15
EVRC	0	3.96
EVRC	1	3.63
EVRC	2	3.82
EVRC	3	3.57

## 2.6. Intermediate Results from PESQ: Frequent AQM

The PESQ standard demands relatively long speech references. This means that each PESQ score will be measured during a long geographical distance (up to 150 meters). However, the PESQ algorithm also provides a means of increasing the geographical resolution. By using results from the last step in the PESQ calculation – see step 10 in Figure 2, page 7 – a PESQ-like score can be obtained for shorter intervals than the whole speech reference. This is done by averaging the split-second disturbance values for each sentence in the speech reference and then transforming them into MOS-LQO scale. The result is two speech quality scores per reference instead of one, which doubles the geographical resolution.

These scores are called Frequent AQM to distinguish them from the regular PESQ scores. The main application of Frequent AQM is pinpointing problems during troubleshooting. An example of the usefulness of Frequent AQM is shown in the map presentation below (Figure 4). In this example the vehicle was driven at a speed of 90 km/h. The measurement period for PESQ is 5.5 seconds, so during that time the vehicle passed through four statistical parcels (each parcel is 50 x 50 m). The PESQ score was attributed to the fourth and last parcel, although the speech quality problem occurred in the first two parcels. The Frequent AQM score, on the other hand, was attributed to the second parcel and thus correctly pinpointed the problem.

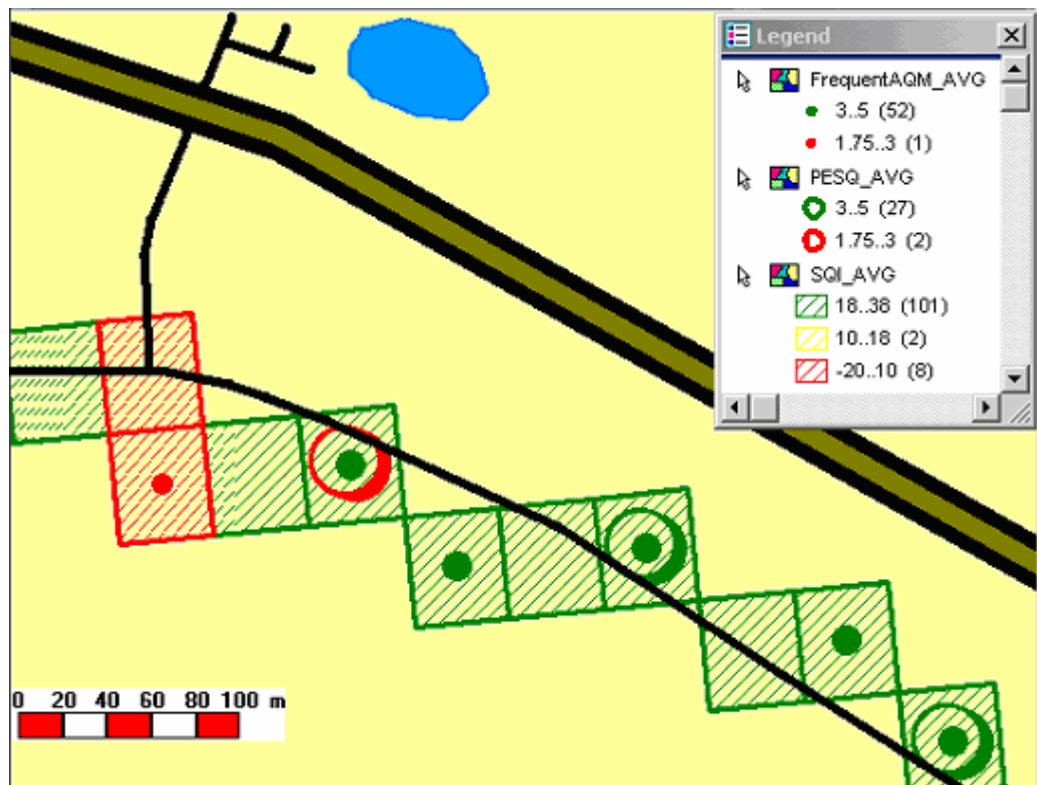


Figure 4 Increasing the geographical resolution with Frequent AQM.

### 2.6.1. How Close Is Frequent AQM to PESQ?

The main difference between PESQ and Frequent AQM, other than the improved geographical resolution, is that Frequent AQM has a slightly higher variance. Typical variations along a route are shown in the graph below.

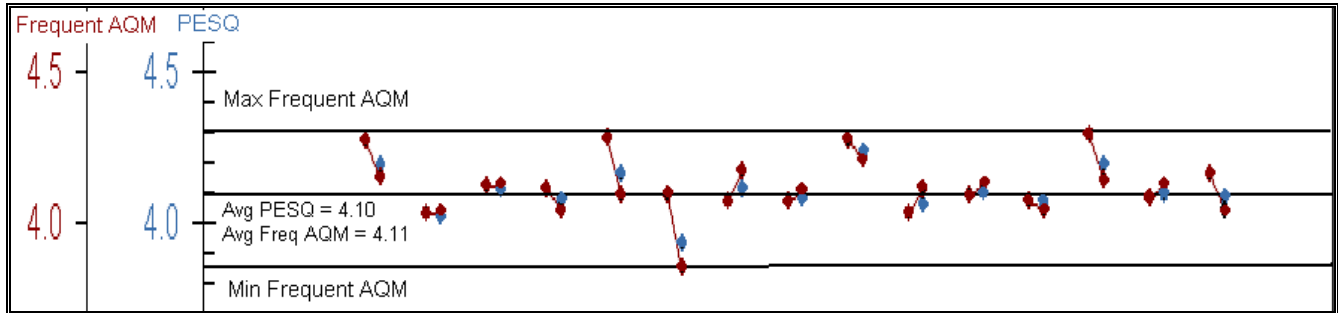


Figure 5 Typical fluctuations for Frequent AQM and PESQ along a route.

The averages of PESQ and Frequent AQM will also differ, but even more slightly. In the table below the average differences are shown between Frequent AQM and PESQ for the GSM EFR and CDMA EVRC speech coders. The average differences are very small (in the order of 0.01 MOS). Still, for benchmarking purposes, when the highest possible accuracy is desired, we recommend using PESQ rather than Frequent AQM. For further comparisons between PESQ and Frequent AQM, see chapter 7.

Speech Coder	Reference	Sentence	Frequent AQM [MOS – LQO]	PESQ [MOS – LQO]	Avg. (Frequent AQM – PESQ)
EFR	0	0	4.30	–	
EFR	0	1	4.17	4.22	0.02
EFR	1	2	4.24	–	
EFR	1	3	4.09	4.16	0.01
EFR	2	4	4.17	–	
EFR	2	5	4.14	4.15	0.001
EFR	3	6	4.16	–	
EFR	3	7	4.11	4.13	0.01
			<b>Avg.: 4.17</b>	<b>Avg.: 4.16</b>	<b>0.01</b>
EVRC	0	0	3.93	–	
EVRC	0	1	4.01	3.96	0.01
EVRC	1	2	3.64	–	
EVRC	1	3	3.66	3.63	0.02
EVRC	2	4	3.80	–	
EVRC	2	5	3.82	3.82	-0.01
EVRC	3	6	3.48	–	
EVRC	3	7	3.71	3.57	0.02
			<b>Avg.: 3.76</b>	<b>Avg.: 3.75</b>	<b>0.01</b>

### 3. PESQ in TEMS Automatic

Audio quality measurements are made in TEMS Automatic during voice calls. The other party in the call can be

- a stationary component known as the Call Generator, connected to the fixed telephone network, or
- a Mobile Receiving Unit (MRU) housed in a mobile phone.

The PESQ algorithm is implemented in the Call Generator and in the MTUs.

See Figure 6.

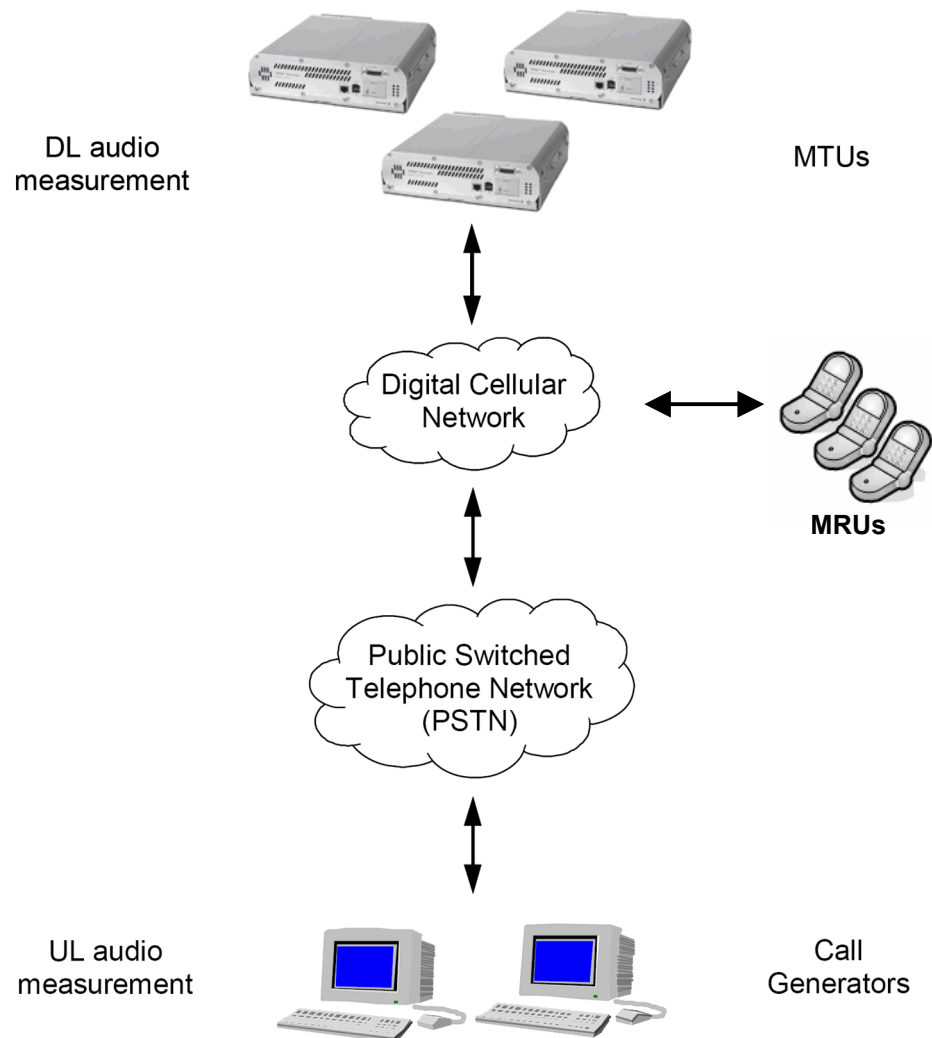


Figure 6 Parts of the TEMS Automatic system that handle AQM. Either Call Generators or MRUs are used (both are shown in the diagram).

### 3.1. MTUs

Each MTU is equipped with an AQM module. The AQM module uses a digital signal processor (DSP) to record and replay speech and to perform the downlink PESQ calculation.

#### 3.1.1. Audio Connection

The usual way for a measurement tool to connect to the mobile's audio is through the analog handsfree port. This port may introduce a degradation which is not in the network, something which may give rise to a difference in estimated quality between uplink and downlink and between different mobile versions. In the TEMS equipment, great care has been taken to adjust volume and mobile audio features so that no degradation will occur.

#### 3.1.2. Monitoring the Audio

Each MTU has a microphone outlet, which makes it possible to listen live to the received speech references. PESQ and Frequent AQM scores are also reported in real time to the MTU trace, so that a user can compare the PESQ scores to his own experience of the quality of the received audio.

It is also possible to record the audio in the MTU, send it up to the TEMS Automatic back-end and listen to the recordings. This is controlled by the Options.ini file stored in the Config directory on the MTU flash disk. The following settings are possible:

AQMLimit=<value>	Recordings with a PESQ score below (i.e. worse than) AQMLimit will be stored. Setting AQMLimit to 4.5 or higher will store everything.
AQMOptions=2	Recordings are stored on the MTU flash disk in the directory "DSP". The recordings are named according to the PESQ score and the time of measurement.
AQMOptions=5	Recordings are sent to the FTP server as .lzz files (same as logfiles).  If the default upload setup is used, the files will be renamed by the ComServer so that they become hard to distinguish from real logfiles. To avoid this, it is best to use a special FTP home directory for all MTU uploads when uploading speech. This directory is set in the Operator Console. The .lzz speech files can be unzipped using Winzip. Note that when using a special FTP home directory, you need to copy logfiles manually to the ComServer.

The recordings are 16-bit linear PCM in Intel format and have a sampling frequency of 8 kHz. For listening to these recordings the program CoolEdit 96 is recommended. CoolEdit also provides advanced audio analysis features which can give extensive insight into speech quality problems.

## **3.2. Call Generator**

The Call Generator is equipped with several telephone boards, which handle the recording and replaying of speech. Several Call Generators can coexist in a TEMS Automatic system, thus allowing an unlimited number of MTUs to make measurement calls at the same time. The PESQ calculation is performed for the Call Generator by the PC processor, using optimized signal processing routines.

If the Call Generator is installed on a multiprocessor machine, it can run PESQ on several processors simultaneously. This is useful in large TEMS Automatic systems.

### **3.2.1. Audio Connection**

The audio connection is established either via the analog PSTN network, using analog phone boards, or via digital ISDN using ISDN boards.

### **3.2.2. Monitoring the Audio**

Recordings with PESQ scores worse than a certain threshold are saved on the Call Generator computer. The user sets the threshold in the Operator Console or in the Call Generator user interface; see the Operator Console and Call Generator documentation. The recordings are A-law quantized with 8 bits and have a sampling frequency of 8 kHz. For listening to these recordings an external program is needed (CoolEdit 96 is recommended). Compare section 3.1.2 (downlink).

## **3.3. Mobile Receiving Unit (MRU)**

The MRU feature is implemented in an ordinary mobile phone, with a new menu where it is possible to set RAT lock and also to switch the MRU function on and off. When the MRU feature is turned on, the phone will be in auto answer mode, which means it will automatically start to play PESQ sentences when it receives a call. The MRU does not report anything to the TEMS Automatic backbone or to TEMS Investigation; all PESQ measurements in a mobile-to-mobile test call originate from the data collecting units. Read more in chapter 5.

## **3.4. Operator Console**

From the Operator Console it is possible to monitor each connected telephone line of a Call Generator and study the PESQ scores for each telephone line in real time. The Operator Console also reports the load on each telephone line and each processor over the last 24 hours.

### **3.5. PESQ Post-processing in TEMS Presentation**

In TEMS Presentation, PESQ and Frequent AQM measurements can be viewed in map, spreadsheet, and line chart windows. Uplink and downlink measurements can be presented together or separately. PESQ downlink scores can be compared with SQI to locate problems in the fixed part of the network. TEMS Presentation also displays extensive statistics on PESQ scores.

PESQ, Frequent AQM, and other downlink measurements from MTUs can also be viewed in TEMS Investigation.

## 4. PESQ in TEMS Investigation

Audio quality measurements are made in TEMS Investigation during voice calls. The other party in the call can be

- a stationary component known as the Call Generator, connected to the fixed telephone network, or
- a Mobile Receiving Unit (MRU) housed in a mobile phone.

The PESQ algorithm is implemented in the Call Generator and in the AQM modules.

See Figure 7.

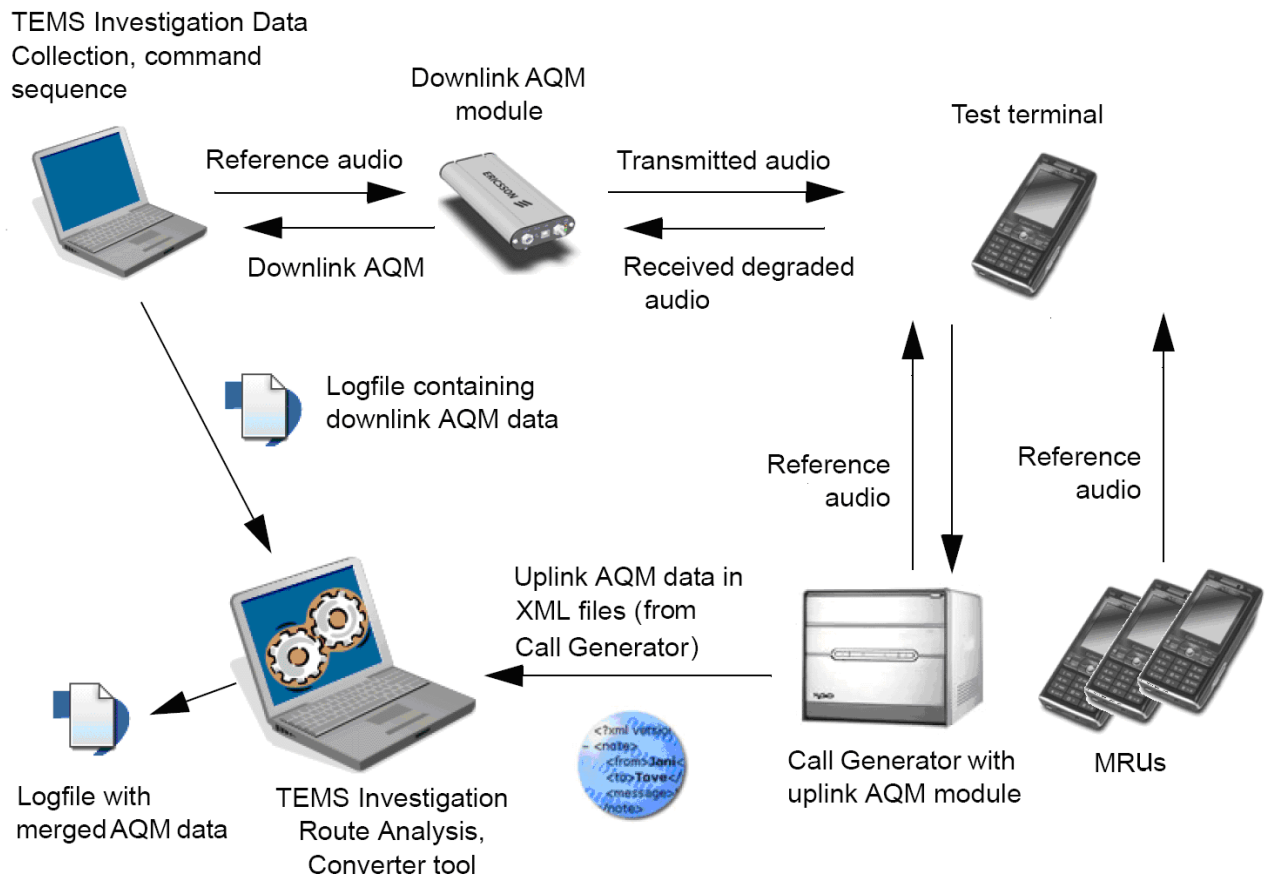


Figure 7 Handling of AQM in TEMS Investigation. Either a Call Generator or MRUs are used (both possibilities are shown in the diagram). The merged logfile will contain positioned DL AQM data and, if applicable, merged UL AQM data from a Call Generator.

## 4.1. TEMS Investigation AQM Modules

TEMS Investigation AQM modules can be either 1) standalone or 2) installed in a TEMS equipment case (up to four AQM modules in one case). Both configurations can be used with TEMS Investigation at the same time.

Each AQM module is equipped with a digital signal processor (DSP) which records and replays speech and performs the PESQ calculation.

### 4.1.1. Audio Connection

The usual way for a measurement tool to connect to the mobile's audio is through the analog handsfree port. This port may introduce a degradation which is not in the network, something which may give rise to a difference in estimated quality between uplink and downlink and between different mobile versions. In the TEMS equipment, great care has been taken to adjust volume and mobile audio features so that no degradation will occur.

### 4.1.2. Monitoring the Audio

Each AQM module has a microphone outlet, which makes it possible to listen live to the received speech references. PESQ and Frequent AQM scores are also reported in real time to the TEMS Investigation application, so that a user can compare the PESQ scores to his own experience of the quality of the received audio.

It is also possible to record the audio in the TEMS Investigation application, so that it can be listened to and analyzed afterwards. This is controlled by the Options.ini file found in the PESQ directory beneath the TEMS Investigation installation directory. The following settings are possible:

AQMLimit=<value>	Recordings with a PESQ score below (i.e. worse than) AQMLimit will be stored. Setting AQMLimit to 4.5 or higher will store everything.
AQMOptions=2	Recordings are stored to the PESQ directory in the TEMS Investigation installation directory. They are named with PESQ value and time of measurement to make them easy to analyze.

The recordings are 16-bit linear PCM in Intel format and have a sampling frequency of 8 kHz. For listening to these recordings the program CoolEdit 96 is recommended. Compare section 3.1.2.

## **4.2. Call Generator**

The Call Generator is equipped with telephone boards, which handle the recording and replaying of speech. Several Call Generators can coexist in a TEMS Investigation PESQ system, thus allowing an unlimited number of UEs to make measurement calls at the same time. The PESQ calculation is performed for the Call Generator by the PC processor, using optimized signal processing routines.

If the Call Generator is installed on a multiprocessor machine, it can run PESQ on several processors simultaneously.

### **4.2.1. Audio Connection**

The audio connection is established either via the analog PSTN network, using analog phone boards, or via digital ISDN using ISDN boards.

### **4.2.2. Monitoring the Audio**

Recordings with PESQ scores worse than a certain threshold are saved on the Call Generator computer. The user sets the threshold in the Call Generator user interface; see the Call Generator documentation. The recordings are A-law quantized with 8 bits and have a sampling frequency of 8 kHz. For listening to these recordings an external program is needed (CoolEdit 96 is recommended). Compare section 3.1.2.

## **4.3. Mobile Receiving Unit (MRU)**

The MRU feature is implemented in an ordinary mobile phone, with a new menu where it is possible to set RAT lock and also to switch the MRU function on and off. When the MRU feature is turned on, the phone will be in auto answer mode, which means it will automatically start to play PESQ sentences when it receives a call. The MRU does not report anything to the TEMS Automatic backbone or to TEMS Investigation; all PESQ measurements in a mobile-to-mobile test call originate from the data collecting units. Read more in chapter 5.

## **4.4. PESQ Post-processing in TEMS Investigation**

The downlink PESQ scores are recorded in the TEMS Investigation logfile, whereas the uplink PESQ scores are stored by the Call Generator. By using the Converter tool in the TEMS Investigation Route Analysis application, you can merge the uplink AQM data into the logfile. The logfile can then be analyzed both in Data Collection and in Route Analysis. The logfile can also be stored into a database, viewed and analyzed in TEMS Presentation.

## 5. Mobile-to-mobile PESQ

TEMS products now support mobile-to-mobile speech quality measurements. This feature requires a node called Mobile Receiving Unit (MRU) in the system. It can be used together with both TEMS Automatic and TEMS Investigation. See Figure 8.

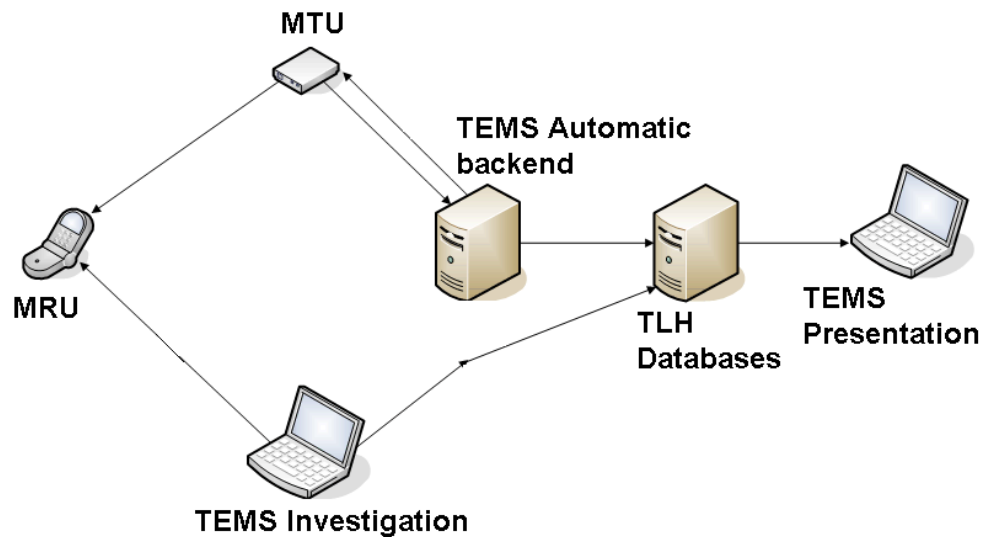


Figure 8 Overview of mobile-to-mobile PESQ.

In mobile-to-mobile PESQ measurements, the MTU (in TEMS Automatic) or the phone device (in TEMS Investigation) makes a call to the MRU, which answers automatically and starts to play PESQ speech sentences. The PESQ algorithm in the receiving AQM module of the MTU for TEMS Automatic or the AQM module for TEMS Investigation calculates and logs the PESQ scores. Those PESQ scores now embrace the entire mobile-to-mobile path, that is, the overall AQM performance including the MRU uplink to the network and downlink to the data collection probe.

### 5.1. Mobile Receiving Unit

The MRU feature is implemented in an ordinary mobile phone, with a new menu where it is possible to set RAT lock and also to switch the MRU function on and off. When the MRU feature is turned on, the phone will be in auto answer mode, which means it will automatically start to play PESQ sentences when it receives a call. The MRU does not report anything to the TEMS Automatic backbone or to TEMS Investigation; all PESQ measurements in a mobile-to-mobile test call originate from the data collecting units.

Some specifics about the MRU:

- The microphone and loudspeaker are muted in MRU mode.

- The MRU disconnects the call after one hour, to avoid hangings in the network or test equipment, and fraud calls.
- The MRU feature in the phone has its own license.

## **5.2. AQM Modules**

### **5.2.1. TEMS Automatic**

Mobile-to-mobile PESQ is initiated by a work order from the Operator Console. The MTU executes the PESQ measurement, which does not contain any DTMF signaling. The MTU continuously records audio and calculates PESQ. The PESQ scores are placed at the correct time and position in the logfile. The audio can be saved to disk on the MTU. No audio is sent to the MRU, and no PESQ is calculated on the MRU side. When using earphones to listen to the recorded speech, only the first reference can be heard, due to limitations in the hardware. In all other respects, the MTU AQM module works in the same way as described in section 3.1.

### **5.2.2. TEMS Investigation**

Mobile-to-mobile PESQ is initiated from a command sequence. The AQM module executes the PESQ measurement, which does not contain any DTMF signaling. The AQM module continuously records audio and calculates PESQ. The Converter tool positions the PESQ and AQM scores at their correct times in the logfile. If the Converter tool is not used, the PESQ scores will be placed 5.5 seconds too late. No audio is sent to the MRU, and no PESQ is calculated on the MRU side. When using earphones to listen to the recorded speech, only the first reference can be heard due to limitations in the hardware. In other respects, the AQM module works in the same way as described in section 4.1.

## **5.3. PESQ Post-processing in TEMS Presentation**

In TEMS Presentation and TEMS Investigation, PESQ and Frequent AQM measurements can be viewed in map, spreadsheet/ AQM monitor and line chart windows. PESQ scores can be compared with SQL to locate problems in the fixed part of the network (when the MRU has been positioned in an error-free radio environment). TEMS Presentation also displays extensive statistics on PESQ scores, and the statistics can be separated on MRU calling number.

## 5.4. How to Use Mobile-to-mobile PESQ Measurement

The mobile-to-mobile PESQ measurement made by an MTU or TI device is affected by both the uplink from the MRU and the downlink to the MTU or TI device. Because radio measurements are made on the downlink only, it is important to ensure that the uplink from the MRU has error-free conditions. Otherwise it is not possible to correlate PESQ with radio data and analyzing a speech quality problem becomes hard. To ensure that the MRU gets error-free conditions, the MRU should be *locked to GSM*, because WCDMA by design aims at non-error-free transmission (usually the target is 1% block error rate for power control reasons). The MRU should also be placed in a good radio environment free from handovers and should be using EFR or AMR full rate speech coding. The radio environment can be checked with a stationary MTU, a TI device, or TEMS Pocket. If possible, calls to a Call Generator should be made to really ensure that the uplink is error-free. In such a call the PESQ average should be within 0.2 of the maximum possible score. Depending on the speech coder and codec rate the maximum possible scores are shown in Table 1 in the TFO column. To summarize, the radio environment for the MRU should have these properties:

- GSM (lock the MRU)
- FER = 0 and RxQual = 0
- No handovers (the serving cell should have a clear RxLev advantage)
- EFR or AMR full rate (heavy network loads should not cause the MRU to use half rate)

These guidelines needs to be abandoned if the test is specifically aimed at investigating other radio technologies or speech codecs, but it is still wise to try to keep the radio environment of the MRU as stable as possible.

### 5.4.1. Impact of Double Speech Coding in Mobile-to-mobile Calls on the PESQ Score

In calls between two mobiles, the speech is usually transported as A-law or Mu-law 64 kbit/s in the core network. This means that the speech is encoded and decoded twice. Networks using the new features Tandem Free Operation (TFO) or Transcoder Free Operation (TrFO) avoid this double coding by transporting the speech in encoded format. In TFO the network commands both mobiles to use the same speech coder, which means the coding and decoding is only performed in the mobiles. This gives three benefits:

- Better voice quality
- Shorter delays
- Less bandwidth needed for transport.

The table below shows how PESQ scores are affected by double coding. In theory, a gain of 0.3 MOS-LQO could be experienced when introducing TFO in a network using EFR. It can also be seen that lower radio quality levels, corresponding to lower AMR rates, are affected worse by the double coding. In a live network using AMR, therefore, the gain from TFO could be higher than 0.3.

Speech Coder	Reference No.	TFO PESQ (MOS-LQO)	Double Coding PESQ (MOS-LQO)	Gain PESQ (MOS-LQO)
EFR	0	4.24	3.97	0.27
EFR	1	4.18	3.75	0.43
EFR	2	4.10	3.72	0.38
EFR	3	4.15	3.83	0.32
<b>EFR</b>	<b>All</b>	<b>4.17</b>	<b>3.82</b>	<b>0.35</b>
AMR 12.2	0	4.19	4.00	0.19
AMR 12.2	1	4.11	3.59	0.51
AMR 12.2	2	4.13	3.83	0.30
AMR 12.2	3	4.06	3.73	0.33
<b>AMR 12.2</b>	<b>All</b>	<b>4.12</b>	<b>3.79</b>	<b>0.33</b>
AMR 7.95	0	3.90	3.43	0.47
AMR 7.95	1	3.64	3.05	0.60
AMR 7.95	2	3.78	3.25	0.52
AMR 7.95	3	3.71	3.16	0.55
<b>AMR 7.95</b>	<b>All</b>	<b>3.76</b>	<b>3.22</b>	<b>0.53</b>
AMR 5.90	0	3.50	3.07	0.43
AMR 5.90	1	3.28	2.73	0.54
AMR 5.90	2	3.35	2.79	0.56
AMR 5.90	3	3.52	2.92	0.60
<b>AMR 5.90</b>	<b>All</b>	<b>3.41</b>	<b>2.88</b>	<b>0.53</b>
AMR 4.75	0	3.23	2.86	0.37
AMR 4.75	1	3.00	2.39	0.60
AMR 4.75	2	3.15	2.53	0.61
AMR 4.75	3	3.25	2.64	0.62
<b>AMR 4.75</b>	<b>All</b>	<b>3.16</b>	<b>2.61</b>	<b>0.55</b>

Table 1: TFO vs. double speech coding.

## 6. PESQ Measurement Procedure

The PESQ measurement procedure is different in mobile-to-mobile PESQ measurement (described in section 6.2) compared with Call Generator to AQM module (and vice versa) PESQ measurement (section 6.1).

### 6.1. Call Generator/AQM Module Measurement Procedure

The measurements are performed in half duplex using a master/slave scheme. Half duplex means that the Call Generator and AQM module alternate between playing and recording speech. Using half duplex is necessary because of the echo canceller in the network. If both the AQM module and Call Generator were active at the same time, clipping would occur. "Master/slave" means that the Call Generator (master) decides the timing of the play/record cycle and that the AQM module (slave) just follows it. The realtime synchronization needed to follow the play/record cycle is described below in section 6.1.1.

Each play/record cycle takes 11 seconds and is repeated over and over throughout the call. During the first 5.5 s, the Call Generator plays speech and the AQM module records it, and in the following 5.5 s the roles are reversed. After each recording three speech quality scores are calculated:

- Frequent AQM for first sentence (2.75 s long)
- Frequent AQM for second sentence (2.75 s)
- PESQ for the entire reference (5.5 s).

This will result in Frequent AQM scores once every 2.75 s and PESQ scores once every 5.5 s, alternately on the uplink and downlink: see below. Buffering of TEMS radio and signaling data ensures that all scores are written to the logfile at the correct time and tagged with the correct geographical position.

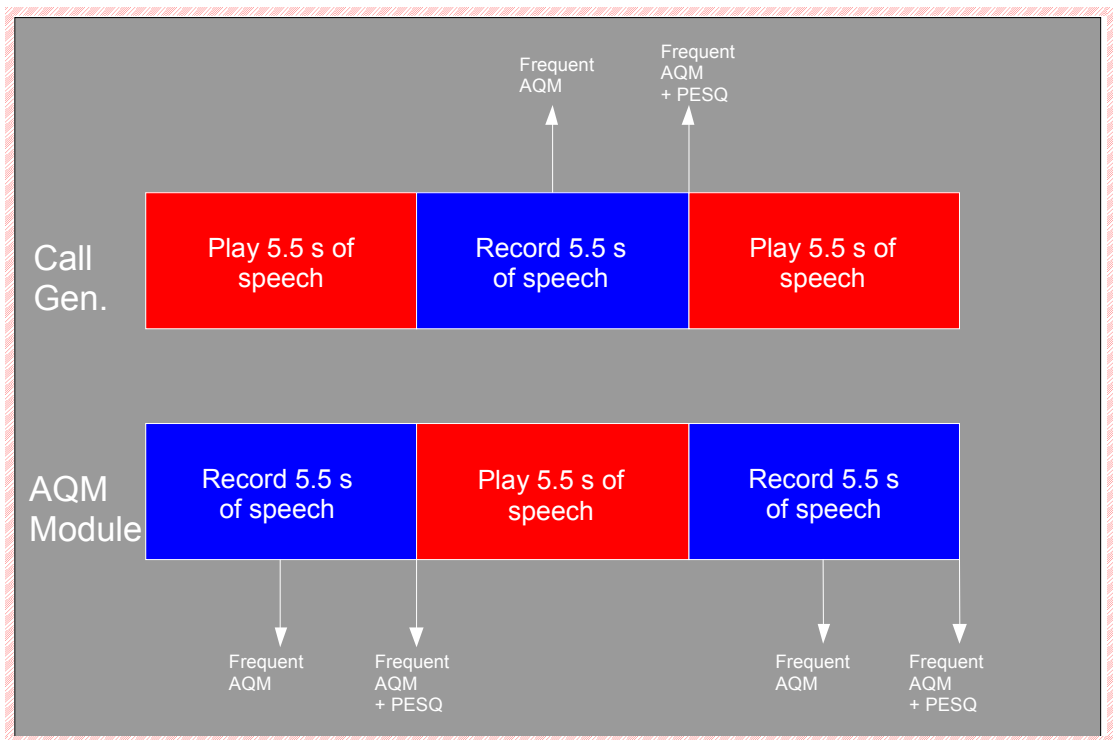


Figure 9 Play/record cycle for PESQ and Frequent AQM calculation.

### 6.1.1. Realtime Synchronization

For each new recording the AQM module has to synchronize itself to the Call Generator in order to adjust to realtime variations in the play/record cycle. The synchronization is however not part of the PESQ standard, and measurement tools from different vendors implement it in different ways. The TEMS method is to synchronize on the speech, thereby avoiding the use of synchronization pulses which would lose valuable measurement time. The synchronization algorithm is an invention by Ericsson. The algorithm searches for small, predefined speech patterns, as exemplified in Figure 10 below.

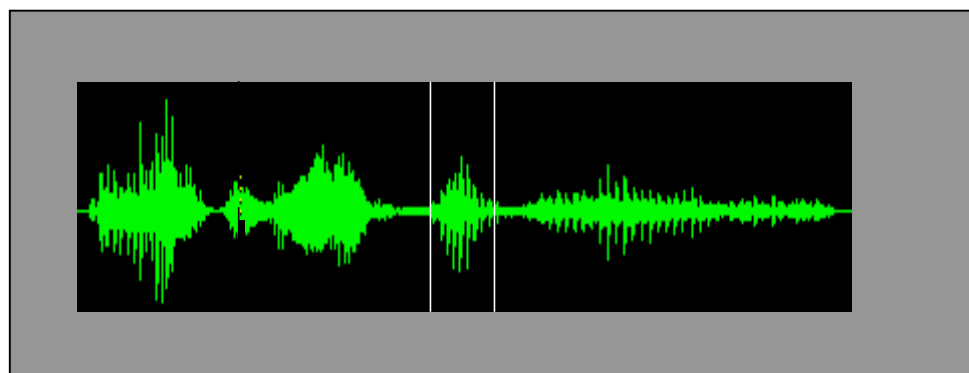


Figure 10 Speech signal with selected synchronization pattern.

### 6.1.2. Details of the Synchronization Algorithm

To increase the calculation efficiency the synchronization algorithm is layered:

- First a fast algorithm searches the complete recording for the sync pattern.
- Then, slower but more accurate algorithms search successively smaller portions around the position indicated by the previous algorithms.

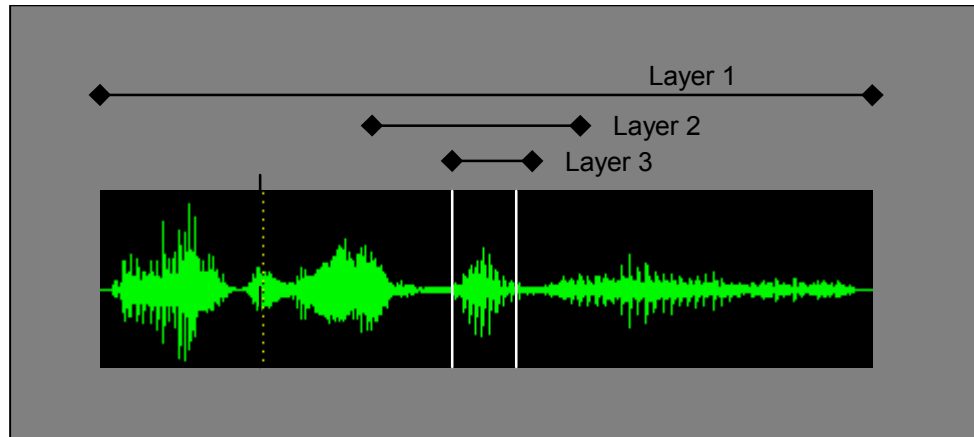


Figure 11 Layering of synchronization algorithm.

To increase the resistance to interference, the synchronization algorithm utilizes the sync from earlier speech sentences. This makes it possible for the algorithm to retain sync even if large parts of a speech sentence have been lost due to muting or corrupted by noise. See Figure 12.

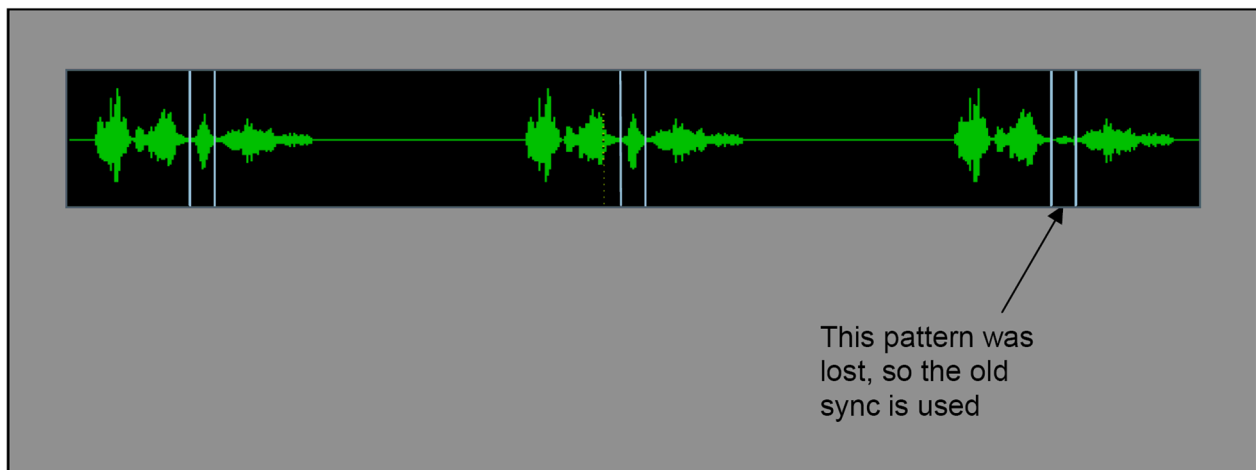


Figure 12 "Fallback" synchronization in case of speech segment deletion.

## 6.2. Mobile-to-mobile Measurement Procedure

The Mobile Receiving Unit (MRU) automatically answers incoming calls and, when the call is established, starts to play the preloaded sentences. Those sentences are then played continuously. The measurements are performed on downlink by the AQM module in the MTU for TEMS Automatic or in the AQM module for TEMS Investigation. For this to work, realtime synchronization is needed (described in section 6.2.1) as well as supervision by the PESQ algorithm (section 6.2.3).

Each play cycle takes 5.5 seconds and is repeated over and over throughout the call. After each recording three speech quality scores are calculated:

- Frequent AQM for first sentence (2.75 s long)
- Frequent AQM for second sentence (2.75 s)
- PESQ for the entire reference (5.5 s).

This will result in Frequent AQM scores once every 2.75 s and PESQ scores once every 5.5 s, for downlink. In the MTUs of the TEMS Automatic system, buffering of TEMS radio and signaling data ensures that all scores are written to the logfile at the correct time and tagged with the correct geographical position . In TEMS Investigation the logfile must be run through the Converter tool to ensure that the PESQ scores are positioned correctly.

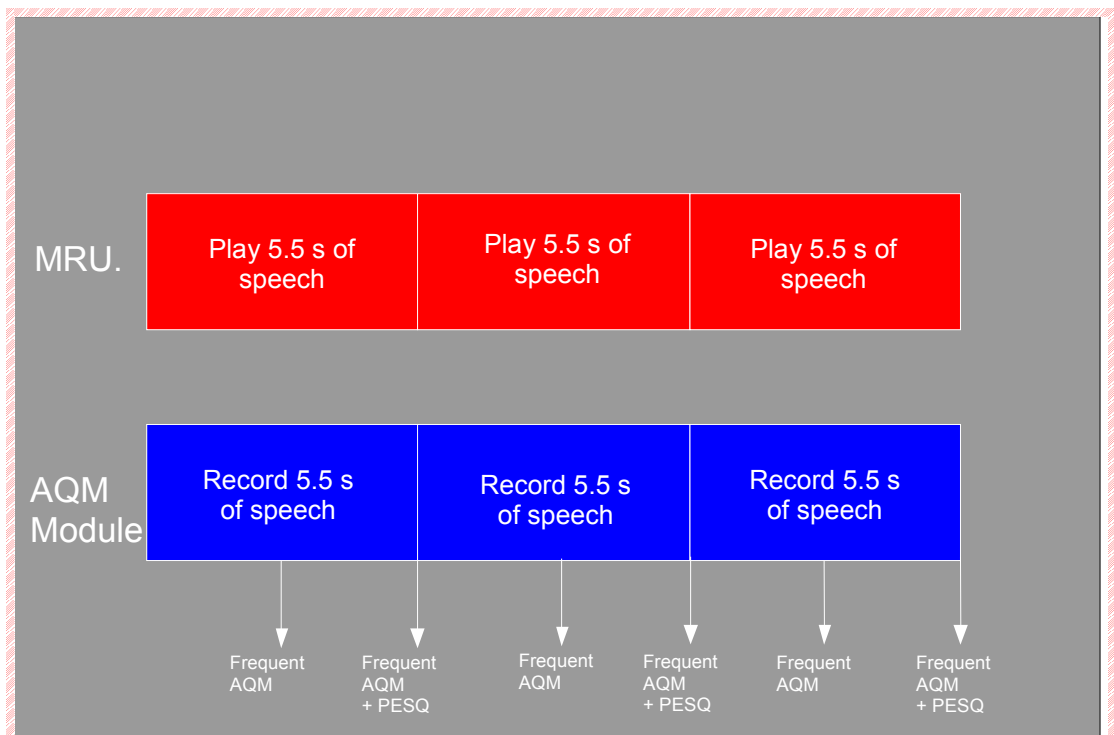


Figure 13 Play/record cycle for PESQ and Frequent AQM calculation: Mobile-to-mobile procedure.

### 6.2.1. Realtime Synchronization

For each new recording the AQM module has to synchronize itself to the MRU in order to adjust to realtime variations in the play/record cycle. The synchronization is however not part of the PESQ standard, and measurement tools from different vendors implement it in different ways. The TEMS method is to synchronize on the speech, thereby avoiding the use of synchronization pulses which would lose valuable measurement time. The synchronization algorithm is an invention by Ericsson. The algorithm searches for small, predefined speech patterns, as exemplified in Figure 14 below.

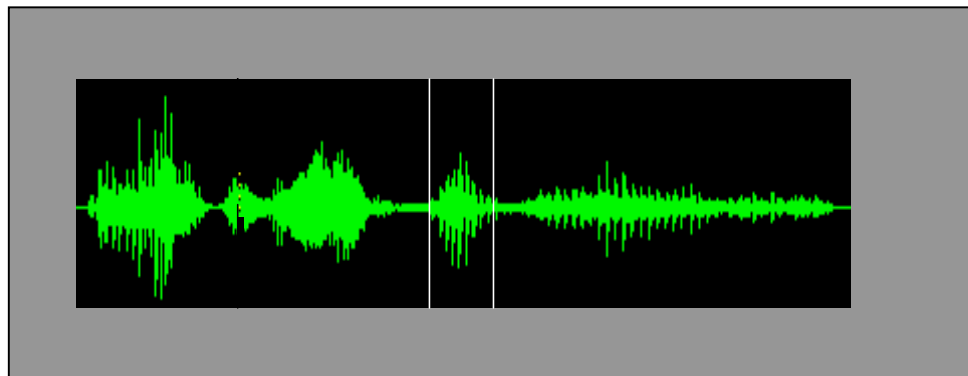


Figure 14 Speech signal with selected synchronization pattern.

### 6.2.2. Details of the Synchronization Algorithm

To increase the calculation efficiency the synchronization algorithm is layered:

- A fast real-time algorithm searches the complete recording for the sync pattern.
- Slower but more accurate algorithms search successively smaller portions around the position indicated by the previous algorithms.

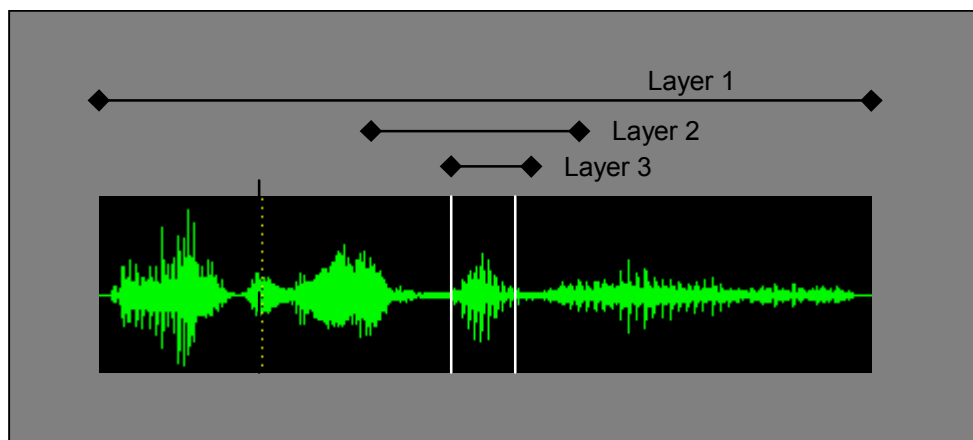


Figure 15 Layering of synchronization algorithm.

### **6.2.3. Mobile-to-mobile Specific Synchronization**

In the beginning of a new recording, the AQM module uses realtime synchronization to synchronize itself to the MRU (see section 6.2.1). When the AQM module is synchronized, it will continue to record without any realtime synchronization applied. The PESQ algorithm will supervise the synchronization, and if the delay exceeds 500 samples the realtime synchronization will be resumed. If realtime synchronization is resumed, there might be a loss of 2–3 PESQ values.

## **6.3. Situations Where No PESQ Score Is Produced**

There are a couple of unusual situations where no PESQ score will be reported:

- When the downlink is silent.
- If the synchronization algorithm fails to find a new sync in four consecutive sentences, no PESQ scores will be given for subsequent recordings until the sync has been found again. The relatively long sentences used for PESQ allow 40 seconds for the radio conditions to improve to the point where the synchronization can again succeed. Most likely the call will fail before the sync fails.
- If a recording contains more than 25% of continuous silence. According to the PESQ standard, the PESQ score is not reliable if the muting exceeds 25%, so we have chosen not to report any score in this case.
- When a call is placed to an MRU, in the unlikely event of exceeding the maximum delay, a new realtime sync will be performed and there might be a loss of 2–3 PESQ scores (see section 6.2.3).

## 7. Verification of PESQ in TEMS Tools

The aim of any speech quality algorithm is to predict the results that would be obtained in a subjective listening test using the same material. In more technical terms, a good speech quality algorithm should have a low prediction error.

- By using the state-of-the-art ITU P.862.1 PESQ standard, we achieve the lowest prediction error that is currently achievable.
- However, since we implemented the PESQ algorithm using optimized hardware and software, it was also necessary to prove the accuracy of that implementation.
- Finally, the audio collection components of the TEMS tools had to be verified.

These three steps are illustrated in Figure 16. Each step is described in more detail in the following sections.

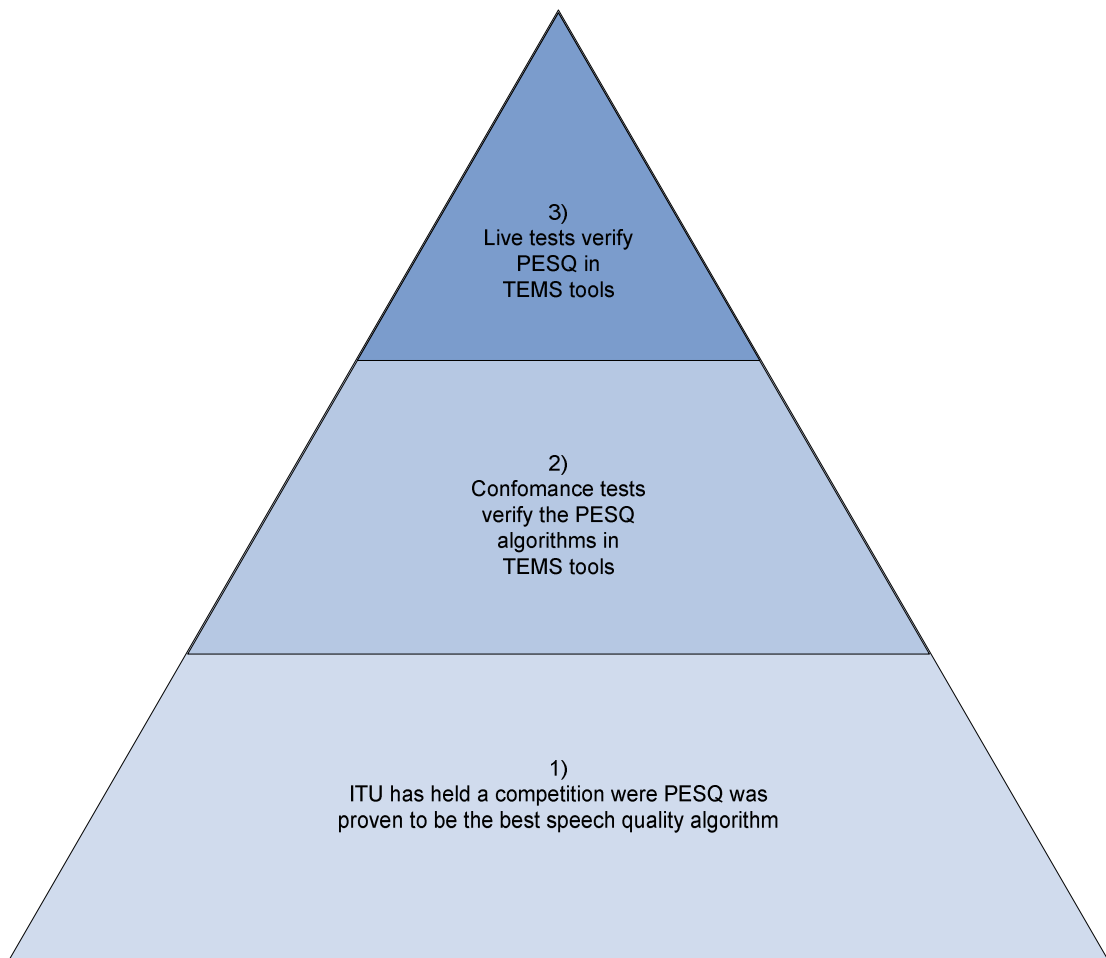


Figure 16 Steps for proving the accuracy of PESQ in TEMS tools.

## 7.1. Step 1: Accuracy of P.862.1

To compare algorithms for speech quality assessment, ITU used a large body of subjectively rated speech data, comprising several languages. These tests led to the choice of PESQ P.862 as the winning algorithm. When the next step was taken in 2003 to develop PESQ P.862.1, even more speech samples were used, including samples from the AMR speech codec. The inclusion of AMR samples in the P.862.1 calibration is especially important because it is the speech coder used in WCDMA. The table below shows the error distribution for P.862 and P.862.1 applied to field-collected data from GSM, CDMA, TDMA, AMPS, and VoIP. P.862.1 shows an improvement over the already very good results from P.862. (The table should be read like this: 70.5% of the P.862.1 scores were within 0.5 MOS of the result from the subjective tests, etc.)

MOS Difference	< 0.25	< 0.5	< 0.75	< 1	< 1.25	< 1.5	< 1.75	< 2
P.862 scores (%)	32.5	66.5	90.8	98.0	99.4	99.9	99.9	100
P.862.1 scores (%)	40.4	70.5	90.3	97.7	99.3	99.7	99.9	100

For these field-collected speech samples (1135 samples in total), the correlation is 93% and the prediction error is 0.42 MOS. Further details on the accuracy of PESQ P.862.1 are found in the Ericsson technical paper "The ITU-T P.862 Standard and Other Speech Quality Metrics and Tools" written by Dr. I. Cotanis, one of the co-developers of P.862.1.

## 7.2. Step 2: Conformance Tests

To enable developers to verify the correctness of PESQ implementations, ITU has provided a large database of audio recordings. The PESQ implementation is to be executed for these recordings, and the results are to be compared with results obtained with an ITU reference implementation. The maximum error must be smaller than 0.05 for all sentences except three; for the latter, the maximum error allowed is 0.1 to 0.5.

Because the Call Generator and the AQM module use different implementations, conformance tests had to be run for the two. The result is shown in Figure 17 below. For both Call Generator and AQM module, only one sentence had an error larger than 0.05 (they were both at 0.06). For most sentences the error was below 0.001. This means that the conformance test was passed for both Call Generator and AQM module.

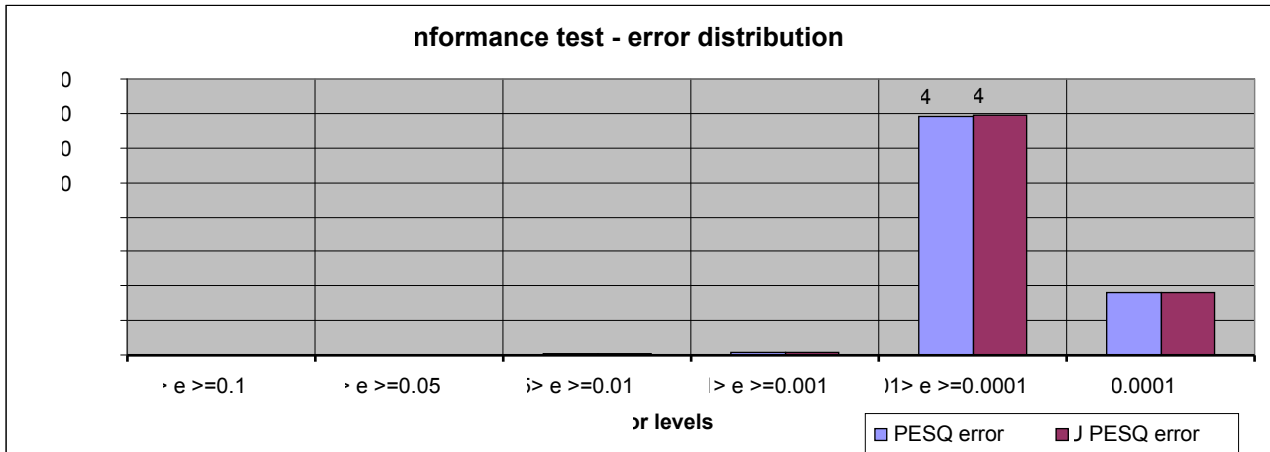


Figure 17 Steps for proving the accuracy of PESQ in TEMS tools.

### 7.3. Step 3: Live Tests

Conducting live tests is the only way to verify an entire measurement tool. There are however two hurdles that one must overcome in order to use live tests for verification:

1. The expected quality under certain conditions must be known.
2. It must be possible to create these conditions in a reliable manner.

Ericsson was capable of satisfying both conditions as follows:

- The expected quality was determined by running the PESQ references through a simulator which mathematically simulates the speech coding, the channel coding, and the actual radio channel. To obtain results that remained stable over time and spanned the most important quality levels, we simulated four of the AMR codec bit rates and the simplest possible radio path: “ideal direct wave channel with no interference”. This means that, apart from making the assumption of an interference-free direct wave channel, the simulation excluded frequency hopping as well as all network features that might affect quality, such as DTX, handover, and background noise reduction. The simulation also did not include degradations arising in the fixed network, such as those caused by A-law coding.
- To produce the “ideal direct wave channel with no interference” condition in a live network, a BTS–BSC–MSC system under Ericsson control was used. It was set up in the same way as the simulation, except that in this case, the fixed network could not be excluded. Very stable conditions were achieved, and each quality level could be measured consistently. The results for uplink and downlink are shown in the table below. As can be seen, the live measurements are very close to the quality expected from the simulations. The uplink is slightly worse than the downlink, but this can be attributed to degradations in the fixed network.

AMR Codec Rate (kbit/s)	PESQ Measured in Live Network, DL	PESQ Measured in Live Network, UL	PESQ Measured in Simulation
12.2	4.11	4.08	4.17
7.95	3.81	3.84	3.84
5.90	3.50	3.50	3.50
4.75	3.21	3.18	3.26

This completes the three steps depicted in Figure 16, and hence we have proved that TEMS tools produce reliable PESQ scores.

## 7.4. Verification of Frequent AQM

With PESQ already verified, the accuracy of Frequent AQM can be proven by comparison with PESQ. For this comparison a field data survey on GSM EFR was used. On the downlink, scores were compared for each quality level, based on RxQual. The average MOS for the total field data showed a very small difference of 0.02 MOS between Frequent AQM and PESQ, which was the expected result; see section 2.6.1. This verifies the AQM module implementation of Frequent AQM.

In poor radio conditions, the variation increased and the difference between PESQ and Frequent AQM grew larger (0.2 MOS in the worst case). For this reason, PESQ should be preferred when benchmarking operators.

RxQual Range	Frequent AQM				PESQ				MOS Diff.
	Avg. RxQual	# of Meas.	MOS	Std. Dev.	Avg. RxQual	# of Meas.	MOS	Std. Dev.	
Overall	1.2	7722	3.93	0.43	1.2	3859	3.91	0.42	0.02
0.0 <= RxQ < 1.0	0.2	4507	4.07	0.16	0.2	2236	4.05	0.14	0.02
1.0 <= RxQ < 2.0	1.4	1255	3.98	0.21	1.5	663	3.95	0.19	0.03
2.0 <= RxQ < 3.0	2.4	900	3.90	0.25	2.4	507	3.85	0.22	0.04
3.0 <= RxQ < 4.0	3.4	519	3.74	0.34	3.4	235	3.65	0.36	0.09
4.0 <= RxQ < 5.0	4.4	292	3.40	0.55	4.4	111	3.25	0.52	0.16
5.0 <= RxQ < 6.0	5.4	154	2.73	0.79	5.4	64	2.55	0.68	0.18
6.0 <= RxQ	6.5	95	1.55	0.40	6.4	43	1.45	0.30	0.10

Downlink Frequent AQM compared to downlink PESQ for field data.

On the uplink the average over all data was compared. Again, the difference was as expected from section 2.6.1. This verifies the Call Generator implementation of Frequent AQM.

	Frequent AQM				PESQ				MOS Diff.
	Avg. RxQual	# of Meas.	MOS	Std. Dev.	Avg. RxQual	# of Meas.	MOS	Std. Dev.	
Total	1.2	7714	3.88	0.29	1.2	3854	3.87	0.28	0.02

Uplink Frequent AQM compared to uplink PESQ for field data.

## 8. Echo Measurements

The purpose of this chapter is to sketch how the echo measuring algorithm works and how it has been implemented in TEMS products.

### 8.1. Overview of Echo Measurement

ITU-T Recommendation G.131 says:

In a two-way telephone conversation a talker sometimes can hear his own voice as a delayed echo, a phenomenon called talker echo. This echo is caused by signal reflections in the transmission path. Most often these occur at 2-wire/4-wire hybrids, sometimes at the listener side as an acoustic feedback from earphone (or loudspeaker) to microphone and possibly also via crosstalk in the handset cord. The perception of echo problems is relative echo delay and echo power, the longer the delay the less power is needed to cause a problem.

The perception of echo depends on both echo delay and echo power. (Diagram in ITU-T Recommendation G.131, figure 1/G.131.)

In national PSTN telephony, the presence of echo is not a problem due to the short (time-wise) signal path – the talker hears his own voice.

In international PSTN telephony, the signal paths will be longer, and thus the presence of echo causes enough of a disturbance to necessitate the introduction of echo limiters/cancelers.

In mobile telephony, the normal 4/2 wire echo delay is compounded by coding and decoding. The cellular network also adds a substantial amount of delay due to the interleaving. This results in a GSM echo delay of at least 160 ms (system constraints). Using this as input in "Talker echo tolerance curves" gives a maximum acceptable talker echo loudness rating of 50 dB. Such good echo attenuation will be achieved only when echo cancellers are used.

Because the echo occurs at the 4/2 wire connection (going from separated UL/DL into a common analog connection) a Call Generator with an analog calling card is needed. Using digital ISDN calling cards or placing calls to an MRU (mobile to mobile) will not produce any echo.

### 8.2. Some Details of the Echo Measurement Algorithm

#### 8.2.1. Calculation of Echo Delay

This calculation is simultaneous with the playback of spoken sentences, which is recorded by the AQM module. The recording is searched for the original sentence, the search being implemented by means of cross-correlation. If the maximum correlation is good enough, the recorded sound is considered to be an echo. The position of the maximum correlation gives the echo delay.

### **8.2.2. Calculation of Echo Attenuation**

Corresponding parts of spoken and recorded sentences are compared with respect to power, and the attenuation is calculated. The resulting attenuation is adjusted for gain settings in the phone and in measurement equipment.

## 9. Silent Call Detection

In some circumstances a call can be set up without actual conversation being possible: all system signaling works, but the audio is not transferred in both directions.

To detect this case an additional check is introduced: if no audio is received, the call is considered to be a silent call.

### 9.1. Some Details of Silent Call Detection

#### 9.1.1. Minimum Call Duration

The call must be long enough that the Call Generator gets to listen to at least four reference sentences. Since the Call Generator sequence is “play, listen, play, ...”, the total time required to detect a silent call is eight times the length of a reference:  $8 \times 5.5 \text{ s} = 44 \text{ s}$ . Since the call setup (guard times, dial time, measurement setup) also takes time, it is wise not to use the shortest possible call duration if the goal is to detect silent calls. The recommended setting is for the call duration to be greater than 90 s.

#### 9.1.2. Silent Call Detection on Uplink

A silent call on the uplink is detected by the Call Generator after it has recorded four consecutive silent speech sentences *at the beginning of the call*. The reason for requiring four silent sentences is that bad radio conditions need to be excluded as the cause of the silence. When the Call Generator has detected a silent call it will end the call (a normal call end). The post-processing will check whether speech was actually sent on the uplink, and a silent call will be reported only if this is true.

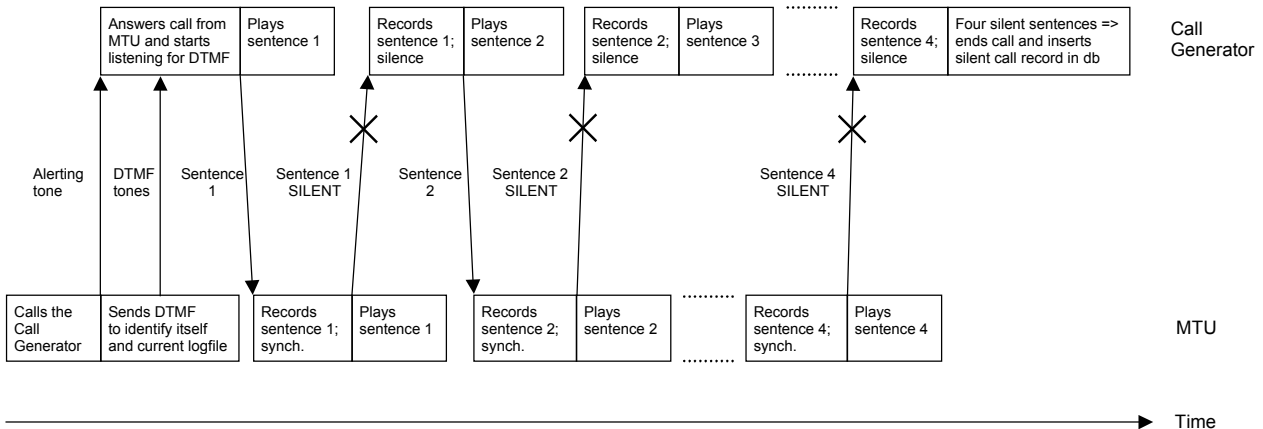


Figure 18 Silent call detection on uplink.

### 9.1.3. Silent Call Detection on Downlink

A silent call on the downlink is detected by the MTU after it has recorded three consecutive silent speech sentences *at the beginning of the call*. Speech is always sent on the downlink, so there is no need to perform additional checks during post-processing: a silent call detected on the downlink will always be reported.

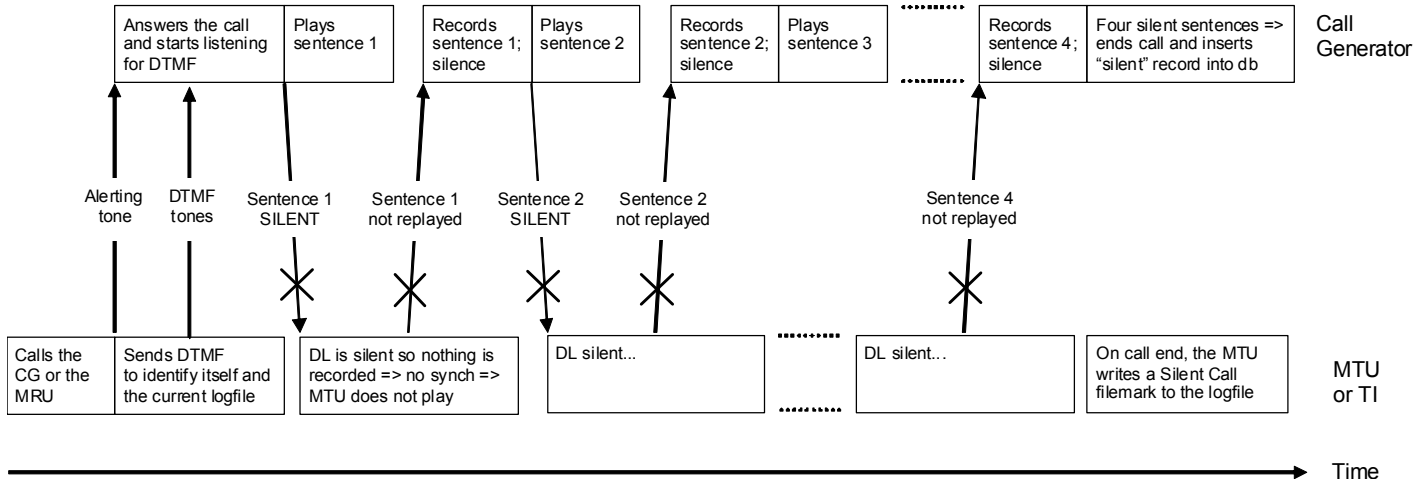


Figure 19 Silent call detection on downlink using a Call Generator. In the MRU case the silent call detection works in a slightly different way but the concept is similar.

### 9.1.4. Call Not Synchronized

A special case is when the MTU cannot synchronize on the downlink speech and therefore does not play any speech on the uplink. The Call Generator will then end the call after recording four silent sentences, as described in section 6.3. No PESQ values will be calculated for such a call, but it will not be regarded as an uplink silent call since the reason for the silence is a synchronization failure.

## 10. Volume Measurements

Volume is measured on the downlink. Volume measurements can be used to check consistency in an operator's network.

**Note 1:** It is not recommended to trim gains on this measurement alone; it is just a way to verify that all parts of the network have the same gain. The volume measurement can also be used to compare gain settings to other operators' settings.

**Note 2:** When using TEMS Investigation, the user has the option of altering the mobile phone volume settings manually. This should be avoided, however, since it will affect the PESQ scores. To set the volume level correctly for specific mobile phones, consult the user documentation accompanying the phones.

### 10.1. Factors Affecting Volume Measurements

#### 10.1.1. Dependency on Sites

Since the equipment that connects the Call Generator affects the volume level, the resulting levels are site-dependent.

#### 10.1.2. Dependency on Mobiles

Not all mobile phones of the same make and model have identical characteristics (as is the case with most hardware). The standard deviation in our evaluation was close to 0.8 dB.

#### 10.1.3. Dependency on Speech Recordings

Different speech sentences give different volume levels. These differences are usually within 1 dB.

#### 10.1.4. Radio Disturbances

Since radio disturbances might give rise to silent segments, they may also affect the volume. Volume readings should be relied on only when the corresponding PESQ values are good.

# 11. Speech Path Delay Measurements

In voice conversations, long delays may cause callers to start talking at the same time and thereby interrupt each other. It can become hard to maintain a meaningful conversation. ITU has investigated how delays affect speech quality; see the table below. If the delay changes during speech it has a severe impact on the speech quality. Delay variations are known as jitter.

Speech Path Delay (Round Trip Time) [ms]	Experienced Quality
< 400	Users very satisfied
400 ... 600	Users satisfied
600 ... 780	Some users dissatisfied
780 ... 1100	Many users dissatisfied

How users experience delay according to ITU Rec. G.114. One way delay has been translated into round trip delay in this table.

## 11.1. Factors Affecting Delay

Delays in the mobile network are caused mainly by speech coding and transport network delay. Speech coding introduces a fixed delay which is always present. The amount of delay depends on the type of call: mobile to PSTN or mobile to mobile. In mobile-to-mobile calls, tandem-free operation (TFO) can be used. In most cases TFO avoids an extra speech coding step, but there are also call scenarios where TFO may cause increased delay. The transport network delay varies depending on the physical length of the connection, whether a satellite connection is used and whether IP transmission is used. It has clearly become more important to monitor delay in modern networks.

## 11.2. Details on the Speech Path Delay Measurement

The TEMS implementation of the speech path delay (SPD) measurement gives the round trip time for the speech. This is the time it takes for speech to travel from the Call Generator (connected to the PSTN or ISDN) to the MTU (mobile in 2G or 3G network) and back to the Call Generator. The SPD measurement is made possible by the real-time sync described in section 6.1.1. For each recording, the real-time sync is updated by the sync from the earlier recording, which means that the accuracy does not become optimal until the second recording in a call. For this reason the SPD is not reported for the first recording in a call. If the Call Generator detects a sync problem, or if PESQ drops below 2.5, SPD is also not reported.

### 11.3. How to Use the Speech Path Delay Measurement

When using MS-to-PSTN calls, the optimal delay is in the range 200–300 ms for simple call scenarios. Even though this is well below the limit of 600 ms when users start to become dissatisfied (according to the ITU G.114 guideline), it should not be interpreted as acceptable for the delay to rise all the way up to this limit. For MS-to-MS calls there is a delay increase of at least 200 ms (without TFO) and more complex call scenarios, with even longer delays, could result in a lot of user calls which are above the 600 ms limit. When monitoring delays it is a good idea to place fixed MTUs at geographical positions which make the call traverse a large part of the network, as in Figure 20 below. Several Call Generator sites can also be used.

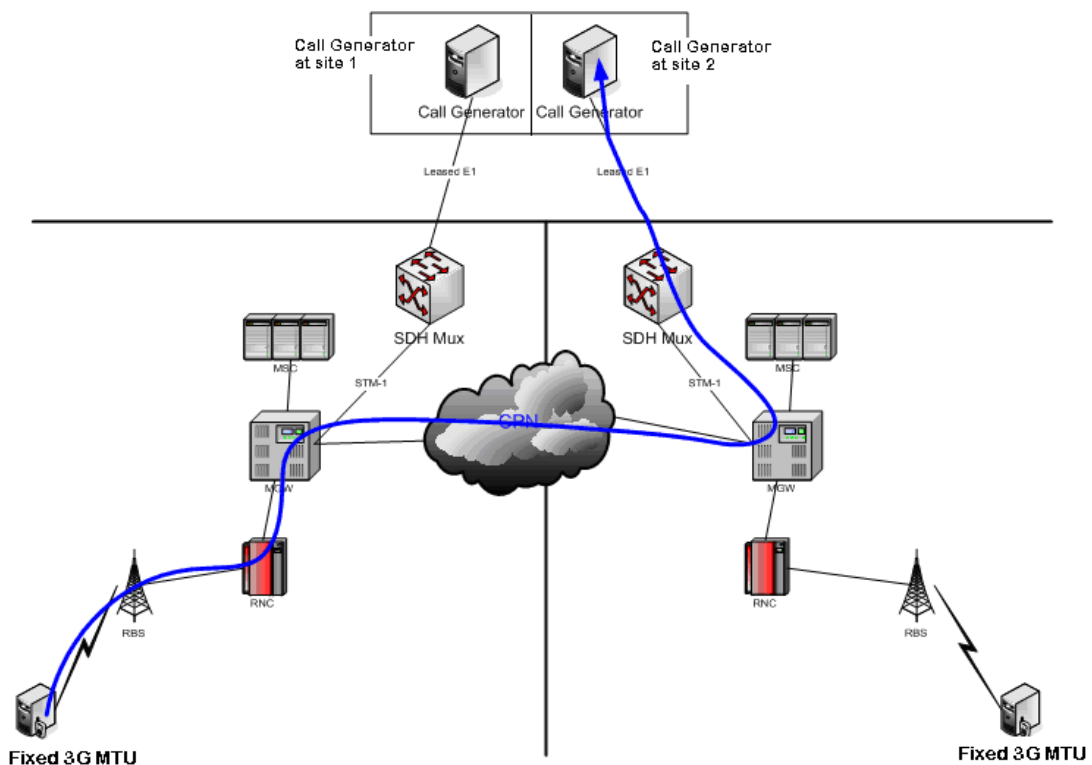


Figure 20 Call scenario in which the call traverses the core network.

The best way to analyze the SPD measurements is to plot a histogram; see Figure 21 below. It is important to monitor the percentage of high SPD values (as a measure of absolute delay) and also to monitor the spread of the SPD values, since an increase in the spread could indicate jitter. For further analysis on a per-call basis, SPD can be plotted in a graph (see Figure 22) or on a map.

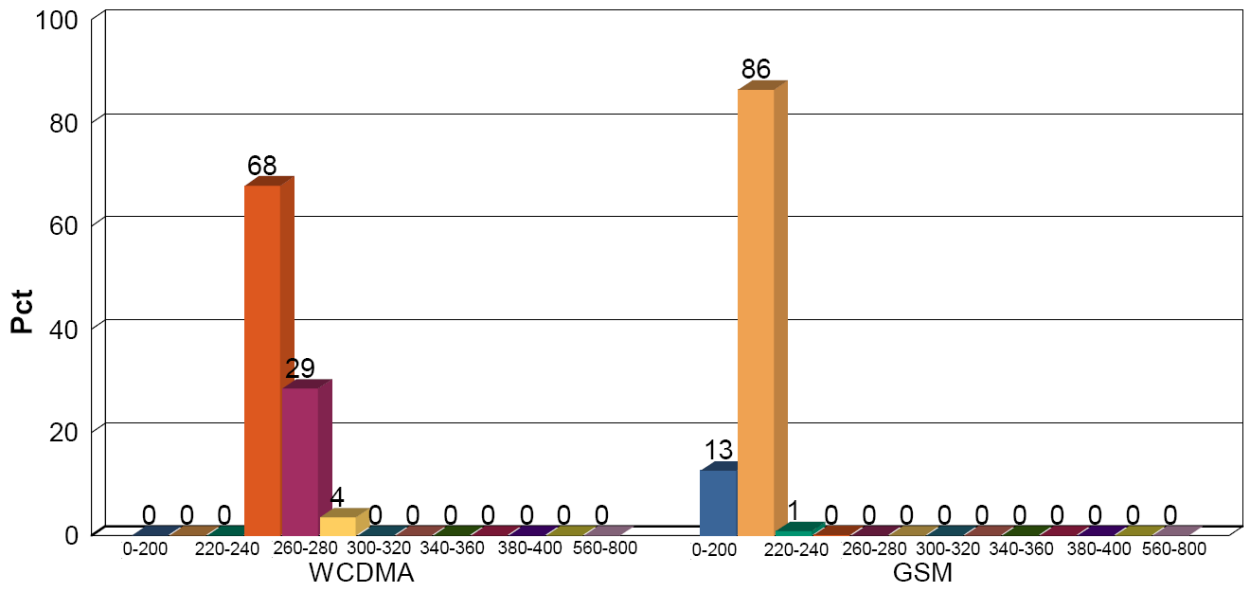


Figure 21 Histogram of speech path delays for MS-PSTN calls on GSM and WCDMA networks.

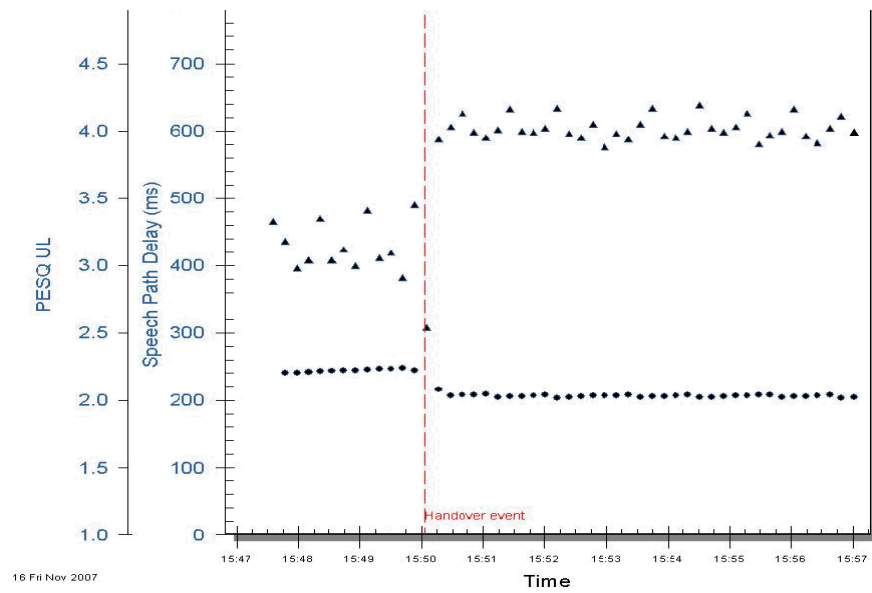


Figure 22 Line chart from TEMS Presentation showing how the speech path delay (dotted curve) changes after a handover.