

# ETSI TR 101 110 v8.0.0 (2000-03)

---

*Technical Report*

## **Digital cellular telecommunications system (Phase 2+); Characterisation, test methods and quality assessment for handsfree Mobile Stations (MSs) (GSM 03.58 version 8.0.0 Release 1999)**

---



**GSM**®  
GLOBAL SYSTEM FOR  
MOBILE COMMUNICATIONS

**ETSI** 

---

Reference

---

RTR/SMG-110358Q7 (aa003i04.PDF)

Keywords

---

Digital cellular telecommunications system,  
Global System for Mobile communications  
(GSM), handsfree, test

**ETSI**

Postal address

---

F-06921 Sophia Antipolis Cedex - FRANCE

Office address

---

650 Route des Lucioles - Sophia Antipolis  
Valbonne - FRANCE  
Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16  
Siret N° 348 623 562 00017 - NAF 742 C  
Association à but non lucratif enregistrée à la  
Sous-Préfecture de Grasse (06) N° 7803/88

Internet

---

secretariat@etsi.fr  
Individual copies of this ETSI deliverable  
can be downloaded from  
<http://www.etsi.org>  
If you find errors in the present document, send your  
comment to: editor@etsi.fr

---

**Copyright Notification**

---

No part may be reproduced except as authorized by written permission.  
The copyright and the foregoing restriction extend to reproduction in all media.

© European Telecommunications Standards Institute 2000.  
All rights reserved.

# Contents

Intellectual Property Rights.....	5
Foreword.....	5
1 Scope.....	6
2 References.....	6
3 Definitions, and abbreviations.....	7
3.1 Definitions.....	7
3.2 Abbreviations.....	7
4 Characteristics, test methods and quality assessment.....	7
4.1 Environmental conditions for a car type handsfree Mobile Station.....	7
4.1.1 Data available on real use environment.....	7
4.1.1.1 Reverberation and echo.....	7
4.1.1.2 Sources and types of noise. Level and Spectra of the noise.....	8
4.1.1.2.1 Noise due to engine, tyres and moving situations.....	8
4.1.1.2.2 Noise due to equipment inside the cars.....	13
4.1.1.2.3 Signal to noise ratio in the car situation.....	13
4.1.1.3 Noise and echo.....	15
4.1.2 Environment for testing handsfree mobile station.....	17
4.1.2.1 Classification of handsfree mobile terminals.....	17
4.1.2.2 Vehicle simulator.....	17
4.1.2.2.1 Cost estimation for a vehicle simulator for handsfree testing.....	17
4.1.2.3 Advisory text for installation of handsfree MS in a vehicle environment.....	18
4.1.2.4 Test environment.....	18
4.1.2.4.1 Anechoic room.....	18
4.1.2.4.2 " Real use " situation (Handsfree in a real car or in a car simulator).....	18
4.1.3 Measurements on a GSM handsfree telephone - Influence of the environment and the test conditions on frequency responses and loudness ratings.....	19
4.2 Delay in handsfree terminals implemented with signal processing techniques.....	22
4.2.1 Signal processing techniques for acoustic echo cancelling and noise reduction.....	22
4.2.2 Examples of delays due to signal processing techniques.....	22
4.2.2.1 Data produced by Matra Communication.....	22
4.2.2.2 Data produced by Ericsson.....	23
4.3 Speech quality assessment.....	24
4.3.1 General - Factors affecting the speech quality of the GSM system and derivatives.....	24
4.3.2 Main Assessment Criteria for Handsfree processing used in GSM mobile environment.....	27
4.3.3 Evaluation Methodology for Full-Duplex Acoustic Echo Controllers developed within the FREETEL-Esprit project.....	28
4.3.3.1 Objective Evaluation procedure.....	29
4.3.3.1.1 Objective Evaluation methodology of AEC devices.....	29
4.3.3.1.2 Test Signals used from the FREETEL-Esprit Database.....	32
4.3.3.2 Adapted objective evaluation procedure to a GSM Handsfree mobile.....	33
4.3.3.2.1 Motivation of an adapted evaluation procedure.....	33
4.3.3.2.3 Proposed Objective-Subjective Evaluation procedure.....	35
4.3.3.2.4 Derived Evaluation schemes for Handsfree mobile telephones.....	35
4.3.4 Subjective tests.....	37
4.3.4.1 Subjective opinion tests.....	37
4.3.4.1.1 Listening opinion tests.....	38
4.3.4.1.2 Conversation Opinion Tests.....	38
4.3.4.1.3 Proposed test method.....	38
4.3.4.2 Subjective tests extracted from "Ultimate test set".....	39

**Annex A: Subjective Tests from UTS**..... 41

**Annex B: Bibliography** ..... 45

B.1 References of TD presented in SMG2/ad hoc 03.50 meetings ..... 45

B.2 References from subclause 4.1 ..... 45

B.3 References from subclause 4.2.1..... 46

B.4 References from subclause 4.3.1..... 46

B.5 References from subclause 4.3.3..... 46

**Annex C: Document change history** ..... 47

History..... 48

---

## Intellectual Property Rights

IPRs essential or potentially essential to the present document may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is publicly available for **ETSI members and non-members**, and can be found in SR 000 314: "*Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards*", which is available **free of charge** from the ETSI Secretariat. Latest updates are available on the ETSI Web server (<http://www.etsi.org/ipr>).

Pursuant to the ETSI IPR Policy, no investigation, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

---

## Foreword

This ETSI Technical Report (TR) has been produced by the Special Mobile Group (SMG).

This specification **gives** guidelines to implement test methods and to evaluate the transmission quality of handsfree Mobile Stations (MSs) within the digital cellular telecommunications system.

The contents of the present document are subject to continuing work within SMG and may change following formal SMG approval. Should SMG modify the contents of the present document, it will then be republished by ETSI with an identifying change of release date and an increase in version number as follows:

Version 8.x.y

where:

- 8 indicates GSM Phase 2+ Release 1999;
- x the second digit is incremented for all other types of changes, i.e. technical enhancements, corrections, updates, etc.;
- y the third digit is incremented when editorial only changes have been incorporated in the specification.

---

# 1 Scope

The present document gives some guidelines to implement test methods and to evaluate the transmission quality of handsfree Mobile Stations (MSs). The normative characteristics and test methods for handsfree Mobile Stations (MSs) are defined in GSM 03.50). The present document gives additional data.

It includes summaries of texts or contributions presented and discussed during the meetings of ad hoc group SMG2/03.50 on environmental conditions, speech processing and quality assessment for handsfree Mobile Station.

The items covered by this report are mainly

- Environmental conditions for handsfree Mobile Stations.
- Speech processing techniques and consequences on delay.
- Speech quality assessment for handsfree implementing acoustic echo cancellation and noise reduction.

---

# 2 References

The following documents contain provisions which, through reference in this text, constitute provisions of the present document.

- References are either specific (identified by date of publication, edition number, version number, etc.) or non-specific.
- For a specific reference, subsequent revisions do not apply.
- For a non-specific reference, the latest version applies.
- A non-specific reference to an ETS shall also be taken to refer to later versions published as an EN with the same number.
- For this Release 1998 document, references to GSM documents are for Release 1998 versions (version 7.x.y).

- [1] I-ETS 300 245-3: "Integrated Services Digital Network (ISDN); Technical characteristics of telephony terminals; Part 3: Pulse Code Modulation (PCM) A-law".
- [2] ITU-T G. 113: "Transmission Impairments Quantisation Distortion".
- [3] ITU-T G.114: "Transmission Impairments, Delay".
- [4] ITU-T Recommendation G.165: "Performances of Network Echo Cancellers".
- [5] ITU-T Recommendation G.167: "Performance of acoustic echo control devices".
- [6] ITU-T G. 711: "64 kbit/s Pulse Code Modulation (PCM)".
- [7] ITU-T G.721: "32 kbit/s Adaptive Differential Pulse Code Modulation (ADPCM)".
- [8] ITU-T G.728: "16 kbit/s Low Delay CELP".
- [9] ITU-T G.729: "8 kbit/s ACELP".
- [10] ITU-T G.731: "Echo".
- [11] ITU-T Recommendation P.50: "Artificial Voices".
- [12] ITU-T Recommendation P.58: "Head and Torso Simulator (HATS) for telephony".
- [13] ITU-T Recommendation P.340: "Transmission Characteristics of Handsfree Telephones".
- [14] ITU-T Handbook On Telephony, Geneva 1987.

- [15] GSM 03.50: "Digital cellular telecommunications system (Phase 2+); Transmission planning aspects of the speech service in the GSM Public Land Mobile Network (PLMN) system".

---

## 3 Definitions, and abbreviations

### 3.1 Definitions

For the purposes of the present document, the following definitions apply:

**Terminal Coupling Loss (TCL):** The frequency dependent coupling loss between the receiving port and sending port of a terminal due to:

- acoustical coupling at the user interface;
- electrical coupling due to crosstalk in the handset cord or within the electrical circuits;
- seismic coupling through the mechanical parts of the terminal.

NOTE 1: The receiving port and the sending port of a digital voice terminal is a 0 dBr point.

NOTE 2: The coupling at the user interface depends on the conditions of use.

**Weighted Terminal Coupling Loss (TCLW):** The weighted Terminal Coupling Loss using the weighting of CCITT Recommendation G.122.

### 3.2 Abbreviations

For the purposes of the present document, the following abbreviations apply:

HATS	Head and Torso Simulator
MRP	Mouth Reference Point
RLR	Receive Loudness Rating
SLR	Send Loudness Rating
TCL	Terminal Coupling Loss
TCLw	Weighted Terminal Coupling Loss

---

## 4 Characteristics, test methods and quality assessment

### 4.1 Environmental conditions for a car type handsfree Mobile Station

#### 4.1.1 Data available on real use environment

##### 4.1.1.1 Reverberation and echo

- a) From Background acoustic noise reduction in mobile telephony (see subclause B.2 reference [1]).

The reverberation times determined from the impulse response inside the cabin are in the order of 25 ms.

- b) From Speech enhancement for mobile telephony (see subclause B.2 reference [2]).

The reverberation time, determined from the impulse response, is around 30ms (it is assumed that the long impulse of the source - a small loudspeaker - may explain why this result is greater than value determined by, reference [1] of subclause B.2.

- c) From Contribution à l'amélioration des performances d'un radiotéléphone mains-libres à commande vocale (see subclause B.2 reference [3]).

The reverberance, in this report, is not defined directly by the reverberation time, but by the part of impulse response energy (from 0 to t), relative to total energy, as a time function.

**Table 1**

Time (ms)	Loss (dB). R25	Loss (dB). 505
10	10.2	9.2
20	22.5	20.2
30	32.5	30.5

#### 4.1.1.2 Sources and types of noise. Level and Spectra of the noise

##### 4.1.1.2.1 Noise due to engine, tyres and moving situations

- a) From Acoustic noise analysis and speech enhancement for mobile radio applications (see subclause B.2 reference [4]).
  - Test conditions.
  - Car: Alfa Romeo Alfetta 200J.
  - Supercardioid directional microphones.
  - Microphone positions: 3 Measurement positions (Front seats, driver and passenger head positions. Middle of rear seat, passenger head position). In these 3 positions, the signals are almost the same.
  - Signal analysis: DFT analysis.

**Table 2**

Testing conditions	Vehicle speed (km/h)	Engine rate (r.p.m.)	Road pavement
a	0	4000	asphalt
b	100	0	asphalt

- Noise
  - in testing condition " a ": a very high peak of energy appears around 120 Hz, at the fundamental frequency of the four strokes four cylinders engine noise. The noise power spectrum decreases hardly between this frequency and 3 kHz.
  - in testing condition " b ": (gear in neutral position, engine kept off), the peak of power in low frequency region disappears and the noise power between 1 kHz and 6 kHz increases considerably relative to condition " a ".
- b) From Background acoustic noise reduction in mobile telephony (see subclause B.2 reference [1])

- Test conditions.
- Measurements were made in a typical mid-size North American car.
- Noise spectrum. see figure 1.



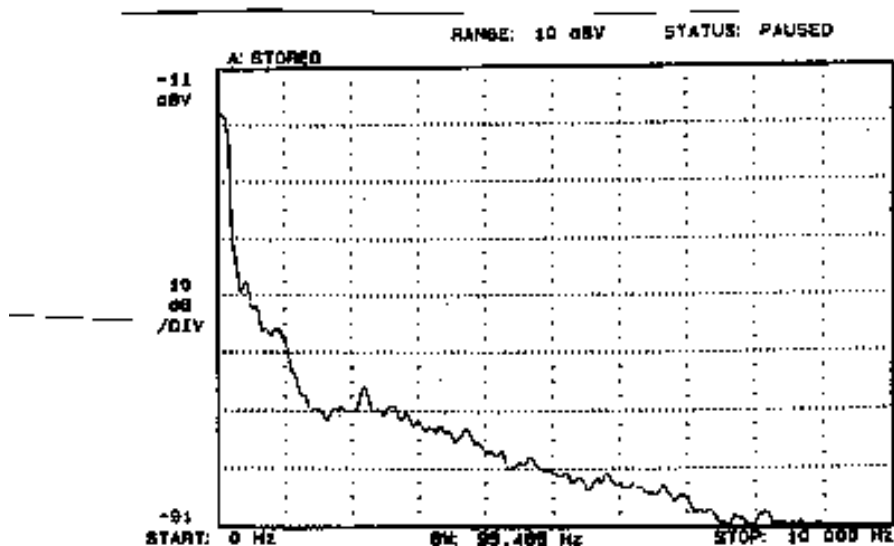


Figure 1: Noise spectrum

The noise power spectrum inside the cabin is mainly located in very low frequencies (below 250 Hz). Comparing the noise outside and inside the cabin, it appears that the car acts as a low pass filter with cutoff frequency around 250 Hz.

Between 250 Hz and about 1.3 kHz, the noise power decreases by about 20 dB.

Above 2 kHz the slope is about -6 dB/octave.

c) From Speech enhancement for mobile telephony (see subclause B.2 reference [2]).

- Test conditions
  - Vehicle: midsized car
  - Speed conditions: idle, 50 km/h, 100 km/h
  - Fan conditions: fan off, fan low, fan high
  - Roads are dry and relatively smooth.
  - Windows are closed.
  - The analysis bandwidth is limited to 4 kHz.
- **Noise.** see figure 2

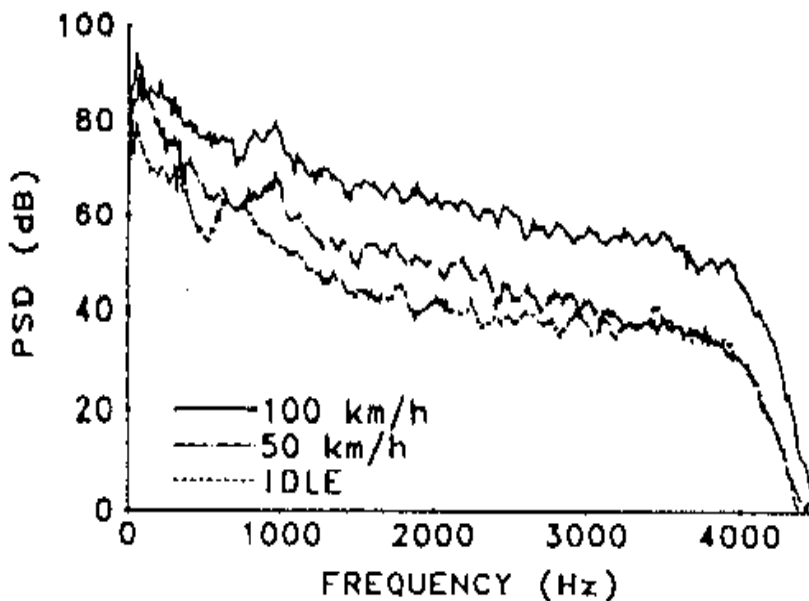


Figure 2: Noise power

With fan off, the maximum of energy is located in low frequencies (below 500 Hz), the peak in the spectrum increasing by about 10 dB from idle to 50 km/h or 100 km/h.

Increasing the car speed increases the noise energy, especially in higher frequencies (more than 10 dB from 50 km/h to 100 km/h).

At high speeds the road noise dominates, while at lower speeds the contribution of the fan to the noise level may become important.

d) From Contribution à l'amélioration des performances d'un radiotelephone mains-libres à commande vocale (see subclause B.2 reference [3])

- Test conditions
  - Cars: Renault 25 and Peugeot 505
  - Omnidirectional microphone
  - The mean value is calculated from about 10 measurements.
- Noise
  - Engine noise

In general, from low frequencies to frequencies in the bandwidth 1 kHz - 1,5 kHz, the level of the spectral components of the engine noise (4 cylinders) decreases by 40 dB. see figure 3 (From Perulli).

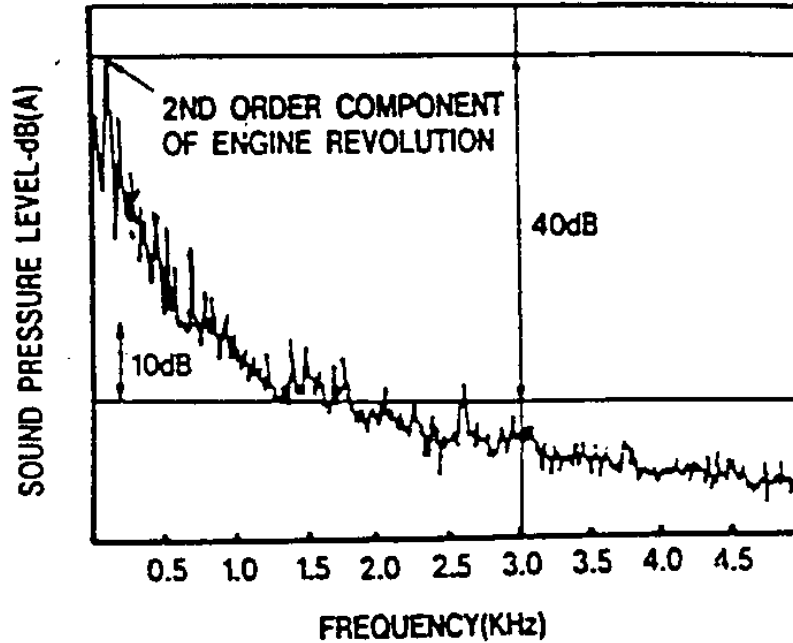


Figure 3: Level of the spectral components of the engine noise

Increasing the engine rate from 2500 r.p.m. to 4000 r.p.m. (that corresponds to speeds about 120 km/h and 150 km/h, depending on the car) increases the noise by about 10 dB.

- Noise due to the engine and to the tyres.

Starting the car: engine noise is dominant.

Urban driving: the two noise sources are equivalent.

On fast roads, tyre noise is preponderant, mainly if the road is wet.

Above 1600 Hz, the type of road has low influence on the noise spectrum.

If the road is wet, above 1 kHz, the level increases relative the dry road condition, and increases with the frequency.

**Supplementary data on the noise generated by the tyres:**

Influence of the granular type of the road (From Perulli)

Table 3

Octave band (Hz)	(1)	(2)	(3)	(4)
31.5	87.5	87	88	89
63	85	86	88	93.5
125	72.5	74	76.5	84
250	71	70	71.5	77.5
500	68	62	64	64
1k	59	58.5	60.5	60
2k	49	48.5	49	47.5
4k	40	39	36	37.5
8k	22	26	23	23

For the conditions (1) to (4) the granular type of the road pavement increases.

- Overall noise

This report gives also results of spectra measured in one car (R25), for two speeds and two types of road. The microphones are located in the middle of the steering wheel and on the sun visor. See figures 4 and 5.

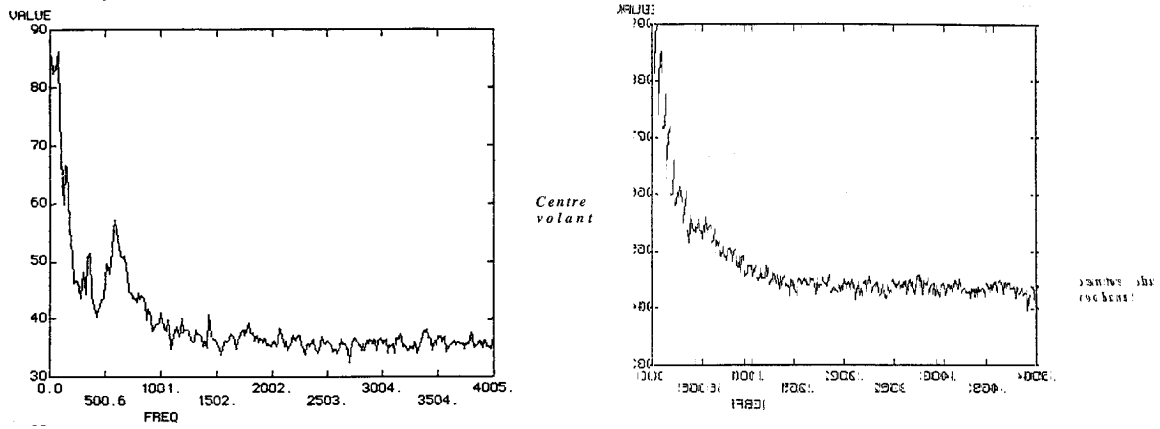


Figure 4:

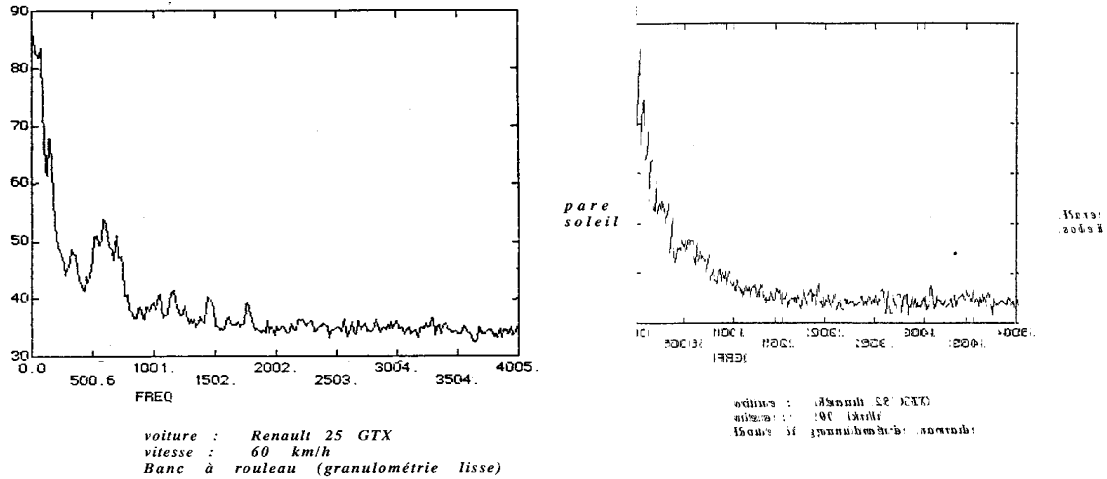


Figure 5:

e) From On the influence of front end processing schemes on the GSM codec behaviour in the context of handsfree radiotelephony (see subclause B.2 reference [5]).

- Test conditions,
- Midsized cars,
- Idle, 70, 100 and 120 km/h,

- Noise

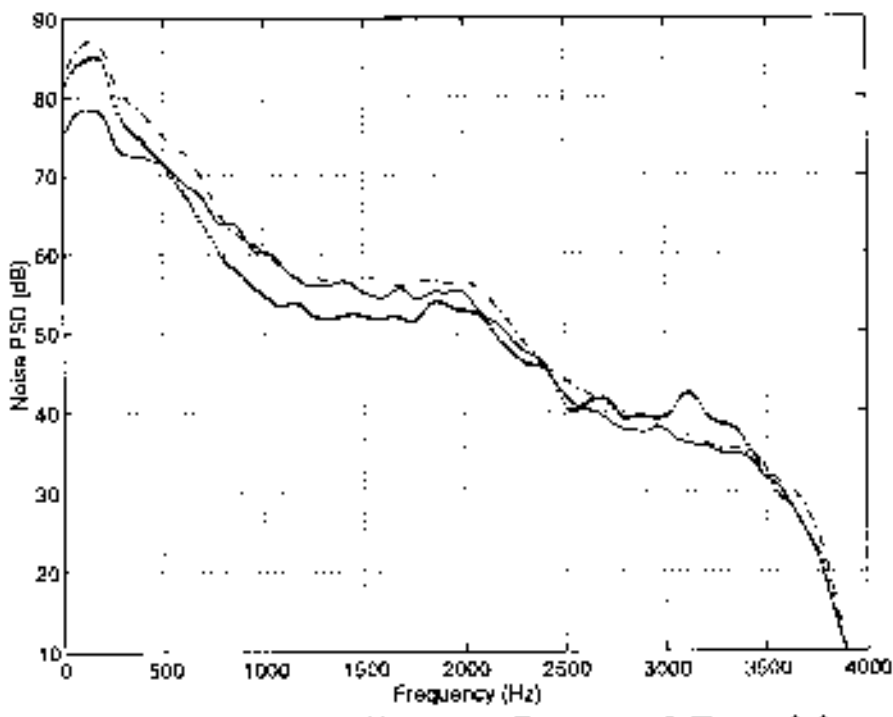


Figure 6: Noise power spectral densities (-) smooth road 100 km/h (..) hard road 100 km/h (-.-) hard road 120 km/h

The noise is dominated by low frequency energy mainly concentrated below 500 Hz.

Between moving on the smooth road and on a hard road, the low frequency energy increases by 12 dB (at a speed of 100 km/h).

The noise caused by the air hitting the windshield appears mainly when the speed is increased. Increasing the car speed from 100 km/h to 120 km/h (on the same road) results in an increase of the noise power by 6 dB, located mainly above 1 kHz.

f) From Noise spectra (see subclause B.2 reference [6])

Supplement n° 13 of the CCITT blue book defines some noise figures for moving and stationary vehicles.

#### 4.1.1.2.2 Noise due to equipment inside the cars

From **Binaural Measurements of loudness as a parameter in the evaluation of sound quality in automobiles** (see subclause B.2 reference [8]).

The results reported are limited to the influence of the heater fan (at different speeds).

The measurements are done by a HATS placed in the car on the driver seat. The two ear simulators are available.

The loudness is calculated according to ISO 532 method B ("Zwicker").

The results are expressed in sones (3 sones being considered as the loudness level of the background noise).

The loudness level due to the fan varies from 3 sones to 6, 11 and 18 sones, as function of the fan speed.

#### 4.1.1.2.3 Signal to noise ratio in the car situation

From **Hands-free and handset mobile telephoning: simulated in situ assessment of telephonic signals and noise using HATS** (see subclause B.2 reference [8]).

Vehicle speed of 100 km/h is simulated on a chassis dynamometer.

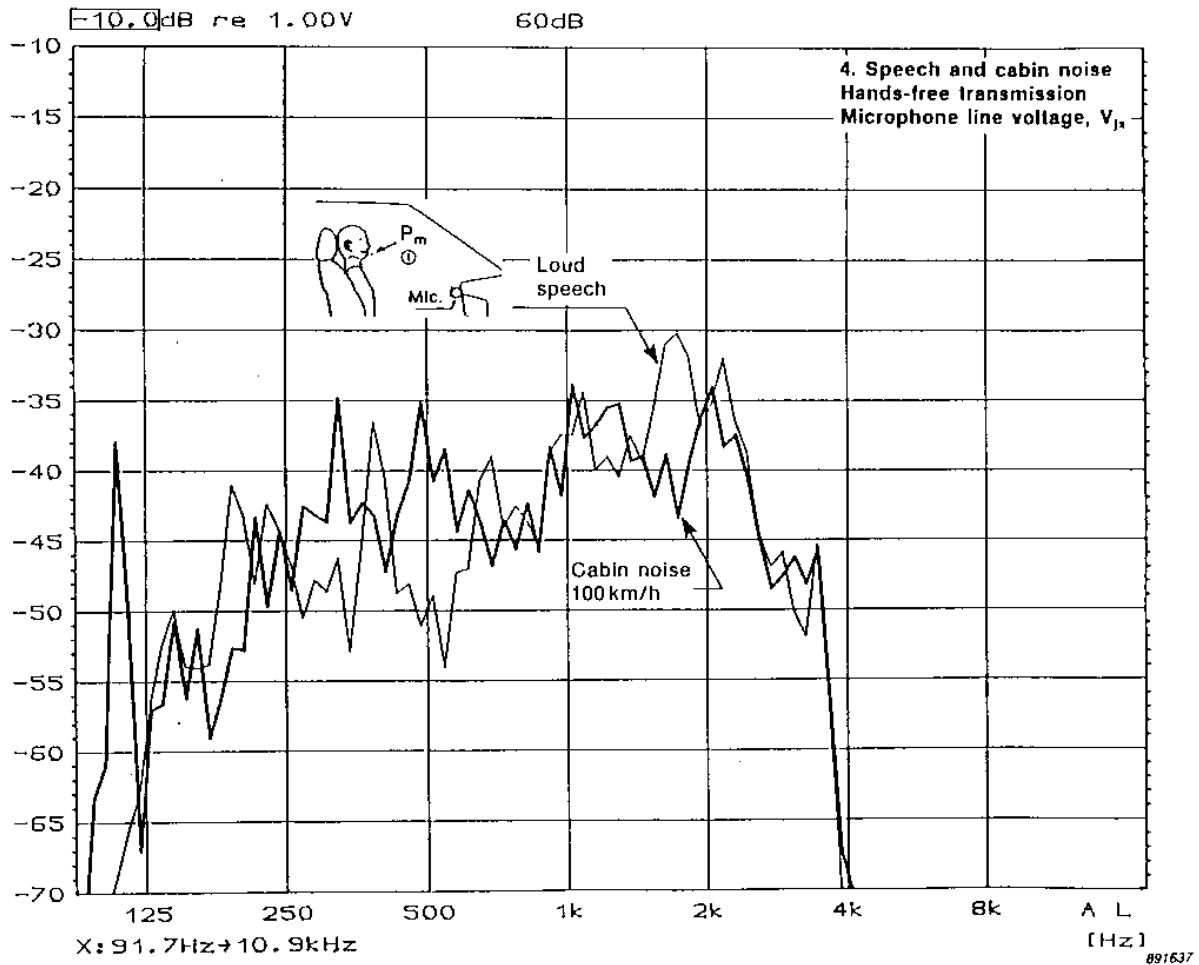
Wind noise not simulated.

Windows and doors closed.

The test is done with a HATS, and through a commercial mobile telephone.

The handsfree microphone and loudspeaker are located at the center console of the dashboard

**Sending**, see figure 7



**Figure 7:**

The speech level produced by HATS is measured at the MRP

Normal speech level: -5 dBPa

Loud speech level: +7dBPa

The noise signal ratio is about +3 dB for loud level, and approximately -10 dB for normal level.

**Receiving**, Figure 8

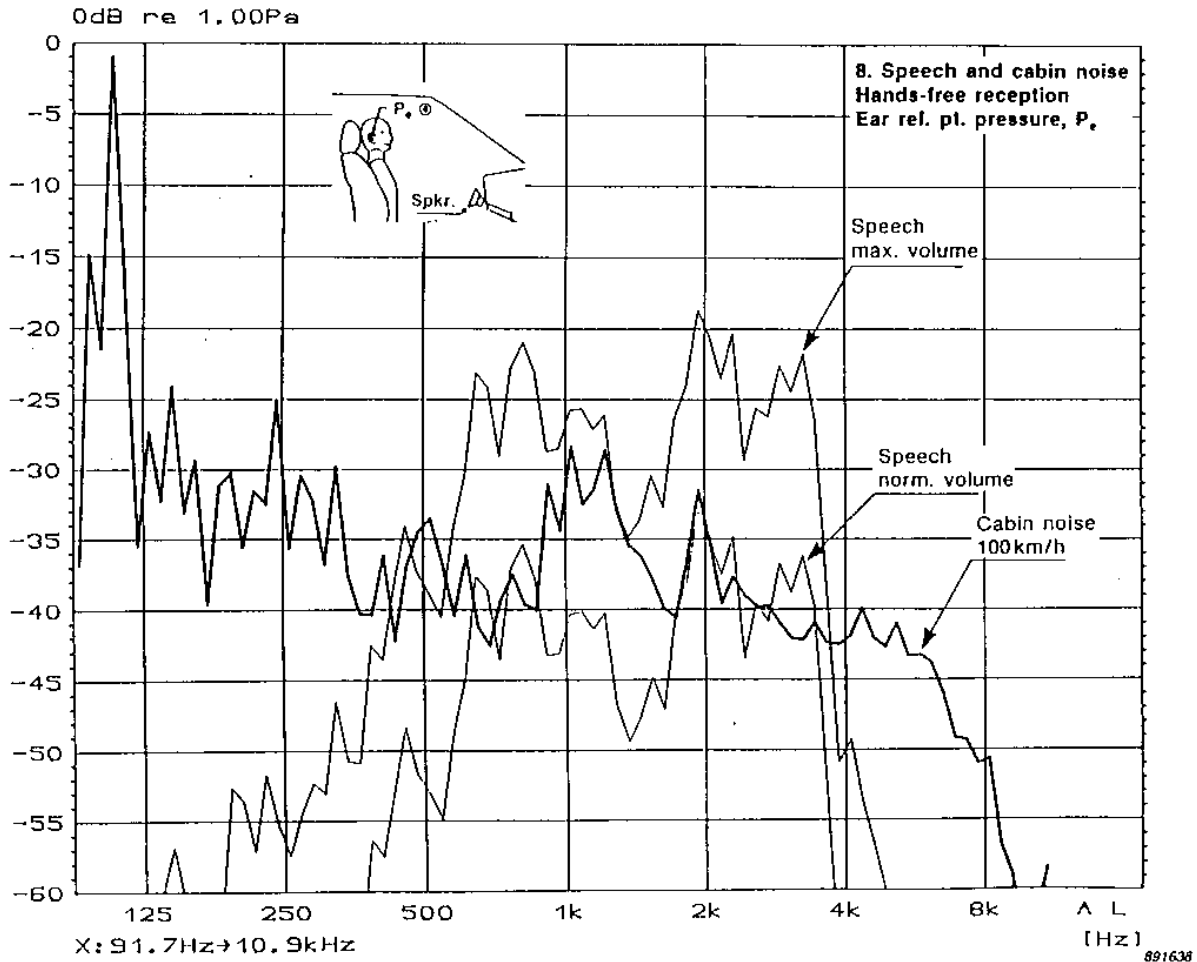


Figure 8

The sound pressure level is measured at the right ERP of HATS, for normal receive volume, and for maximum volume.

The noise signal ratio is about +9,5 dB for maximum volume setting, and approximately -4,8 dB for nominal volume setting.

#### 4.1.1.3 Noise and echo

a) From Subjective evaluation of quality of communications in car hands-free radiotelephone situation (see subclause B.2 reference [9]).

Two simulated handsfree sets are placed in a car, as defined below:

Table 4

Test conditions	Microphone	Loudspeaker
RT1	Sun visor	Central position on dashboard
RT2	Seat belt	Head rest

The subjective test is made by distant user using a handset.

A HATS is placed in the car, on the driver seat.

Noise in the car is generated by loudspeakers at 3 different levels:

B0 (idle situation, engine off),

B1 (90 km/h),

B2 (130 km/h).

3 delay conditions (round trip delay: 0 ms, 60 ms, 180 ms) are available.

3 values for ERLE (S+ 2 dB, S+5 dB, S+11 dB) are introduced in the echo path , for each Bi condition.

The signal to noise ratio is:

about 11 dB for B1

about 6 dB for B2

ORL for RT1 condition

**Table 5**

Handsfree to handset	Handset to Handsfree		
	B0	B1	B2
9.8 dB	4.6 dB*	-5.4 dB*	-10.5 dB*
* These levels were adjusted before the experiments by operators.			

An attenuation of 25 dB or 28 dB seems necessary to obtain good or excellent quality level (DMOS > 4) in B0 situation, for the 3 delay conditions.

A minimum attenuation of 15 dB seems to be necessary at 90 km/h (B1) to ensure a satisfactory quality level (echo DMOS > 3.5), with round trip delay of 60 ms and 180 ms.

A minimum attenuation of 12 dB seems to be necessary at 130 km/h (B2) to ensure an acceptable quality level (echo DMOS > 3), with round trip delay of 60 ms and 180 ms.

b) From Subjective evaluation of quality of communications in car hands-free radiotelephone context (see subclause B.2 reference [10]).

This experiment differs from those described in reference [9] of subclause B.2, mainly on:

- conversation situation instead of a listening test.
- the "driver" has two tasks: he uses a driving simulator and he participates to the conversation.

2 delay conditions (round trip delay: 0 ms, 180 ms).

3 values for ERLE (S+2 dB, S+5 dB, S+8 dB) are introduced in the echo path , for each Bi condition.

Signal to noise ratio is:

about 11 dB for B1

about 6 dB for B2

It appears that noise is more annoying than echo for the distant user (handset).

The detectability threshold of the echo where obtained as:

**Table 6**

<b>B0</b>	<b>28 dB</b>
B1	19 dB
B2	13 dB



## 4.1.2 Environment for testing handsfree mobile station

### 4.1.2.1 Classification of handsfree mobile terminals

Three categories of handsfree equipment:

- The lap-top pc or dedicated desk top handsfree terminal. It could be tested using the desk top style tests listed in I-ETS 300 245-3.
- Integrated in car solutions, where the electronics of the GSM handsfree terminal are inseparable from the vehicle in which it is installed, will be tested with the car in which they are installed.
- Non-vehicle specific handsfree terminals, that should be tested a vehicle simulator.

A handsfree terminal, designed for retro-fitting to any vehicle, would need to be tested in a representative environment (e.g. the vehicle simulator). In addition, with each terminal potentially requiring different mounting positions, any "standard" test car owned by a test house would soon be in a poor state of repair due to the constant alterations to the interior to fit new handsfree terminals for testing. A difficulty with a vehicle simulator is that it needs to be built and verified before it can be used.

Since any manufacturer developing a non-vehicle specific handsfree terminal will have mounted it in a suitable vehicle for development testing, and that the acoustic properties of a given vehicle are likely to be just as representative of that class of vehicle as the acoustic properties of the vehicle simulator, it is proposed that the non-vehicle specific handsfree terminals are tested in the manufacturer's development testing vehicle.

Certain criteria should still be set for the vehicle such that the tests are meaningful across a selection of terminals. These criteria should be that the car provided by the manufacturer should be a medium sized family. Any non-mobile terminal connected to noise reduction systems in the car should be disabled. These criteria are designed to ensure that the test houses will always have test accommodation able to deal with the cars provided by the manufacturers and that the ambient noise reduction test, evaluates the mobile not the car.

So, to overcome the funding and timescale issues surrounding the production of the vehicle simulator, it is proposed that non-vehicle specific handsfree terminals are tested in a vehicle of the manufacturers choice, supplied to the test house with the handsfree equipment pre-installed in the desired mounting position by the manufacturer. This means that the test facilities required by the test house will be the same as for vehicle integrated handsfree terminals, i.e. a garage with suitable noise environment generation. The cost should be minimal for the manufacturers as they should have carried out development testing on their terminals in a vehicle prior to seeking type approval and hence they can use the development test vehicle.

### 4.1.2.2 Vehicle simulator

#### 4.1.2.2.1 Cost estimation for a vehicle simulator for handsfree testing

The following cost estimation is based on a simplified vehicle simulator with "wooden" windows and a very simplified interior.

In order to build up and validate a vehicle simulator the following points are necessary:

1) Construction and build up of a prototype including:

means for fixing microphone, loudspeaker and an artificial head in a defined, reproducible position, loudspeaker for room noise insertion.

2) Validation of the simulator including:

measurements in 4 "representative" medium sized cars for two different sets (implementations).

The costs of such a development would be around 90.000,- ECU.

The price for a simulator could be in the range of 17.500,- ECU.

### Driving simulation for subjective tests

Driving corresponds to specific physical and mental activities. It could be useful to perform some tests (e.g. conversation tests) in such conditions, using a driving simulator.

#### 4.1.2.3 Advisory text for installation of handsfree MS in a vehicle environment

"The hands free specification within ETS 300 903 (GSM 03.50) is designed to provide a basic level of performance and to avoid adverse interactions with other networks. Testing is carried out using a vehicle simulator to standardise the assessment environment. It is designed to be representative of the vehicle environment rather than mirrors the exact properties of particular vehicles. Manufacturers, submitting a handsfree MS to a test house, should assume the standard acoustic transducer positions or state alternative the mounting positions of the transducers".

"In a real vehicle, care should be taken to allow for the acoustic properties of that vehicle and the likely acoustic environment. It is important that the best possible coupling between the microphone and the MS user is achieved. Hence, the microphone should be directional and mounted as close to the users mouth as possible. The loudspeakers should be mounted in such a way that the maximum received signal is directed at the user, rather than dissipated by the various obstacles in the vehicle, such as the seats."

"Proper consideration for the noise environment and the direct coupling between the microphone and loudspeakers is necessary. Excessive noise coupled into the microphone can mask the MS users sent speech and potentially affects the operation of DTX. The vehicle noise environment can potentially mask the received speech unless sufficient volume is provided. However, direct coupling between the transducers can cause annoying echo to be heard by the far end user."

"Primary factors affecting the coupling between the loudspeaker(s) and microphone(s) include:

- Directionality of the microphone(s)
- Directionality of the loudspeaker(s)
- Location of the transducer in relation to reflecting surfaces such as the windows and windscreen".

#### 4.1.2.4 Test environment

##### 4.1.2.4.1 Anechoic room

For the lap-top PC or dedicated desk top handsfree terminals, the test conditions shall be in conformance with I-ETS 300 245-3.

For vehicle handsfree terminals, some tests could be done in an anechoic environment.

These tests could permit to check the transducers or/and electroacoustic equipment performance, like frequency response, harmonic distortion.

The results shall be adapted to take into account the differences between the anechoic conditions and the real use environment.

**NOTE:** If the transducers have special connectors permitting to be unplugged from the handsfree terminal, the test could be performed on the transducers alone, the impedance's of the amplifiers used in the test equipment being adapted according to the manufacturer instructions.

This test only qualifies the transducers and does not give enough information on the quality of the complete handsfree terminal (in particular these tests do not take into account the speech processing, coding,...).

##### 4.1.2.4.2 " Real use " situation (Handsfree in a real car or in a car simulator)

- Noisy environment

The table gives a set of reference conditions for noisy environment.

Table 7

Speed (km/h)	Inside noise	Outside noise	Road surface
0		Road noise (car and lorry traffic) <i>Non stationary noise</i>	
0		Street noise (people talking outside the car) <i>Non stationary noise</i>	
0	Noise of the car engine <i>Stationary noise</i>		
0	Noise of the fan <i>Stationary noise</i>		
0	Speech or/and music <i>Non stationary noise</i>		
50 (or 70)	Noise of the car engine <i>Stationary noise</i>		Smooth, dry
50 (or 70)	Noise of the car engine <i>Stationary noise</i>		Rough, dry
50 (or 70)	Noise of the car engine <i>Stationary noise</i>	Rain falling down on the car	Wet
50 (or 70)	Noise of the car engine <i>Stationary noise</i>	Road and street noise <i>Non stationary noise</i>	Smooth, dry
120 (or 130)	Noise of the car engine <i>Stationary noise</i>		Smooth, dry
120 (or 130)	Noise of the car engine <i>Stationary noise</i>		Rough, dry
120 (or 130)	Noise of the car engine <i>Stationary noise</i>	Rain falling down on the car	Wet
120 (or 130)	Noise of the car engine <i>Stationary noise</i>	Road and street noise <i>Non stationary noise</i>	Smooth, dry

For all these noises, different samples shall be needed to avoid the behaviour of the speech processing systems be optimised on the test references.

- Reverberant environment

A simulated reverberant environment, based on data available in different laboratories from real impulse responses shall be produced in the reference test environment. This environment is intended to simulate the reverberant characteristics of the car interior.

### 4.1.3 Measurements on a GSM handsfree telephone - Influence of the environment and the test conditions on frequency responses and loudness ratings

This subclause presents some experimental results.

Measurements of frequency response curves have been made on a GSM telephone terminal with hands-free function, mounted in a car and also according to the principles in ITU-T Recommendation. P.340 (tests performed under free-field conditions). Loudness Ratings have also been calculated.

The measurements only represent one example out of many possible implementations, and cannot be the basis for any general conclusions.

#### TEST SET-UP

The loudspeaker was mounted on the dashboard of the car. Two types of microphones were used. One was a stick-on microphone intended to rest on the chest of the user, the other was a goose-neck microphone with the possibility to adjust the microphone to a perfect speaking position.

For the measurements at sending an artificial mouth was used. The chest microphone was hanging in free air in the car 20 cm in front of the mouth (on axis). The measurement was repeated at the same distance but in an anechoic room.

The sensitivity of the goose-neck microphone was only measured in an anechoic room.

The measurements at receiving was made with two different loudspeaker positions in the car, both at a distance of 85 cm from the measuring microphone, corresponding to the driver position. The measurements were repeated in an anechoic room according to the principles in ITU-T Recommendation. P.340.

The test set-up is shown in figure 9.

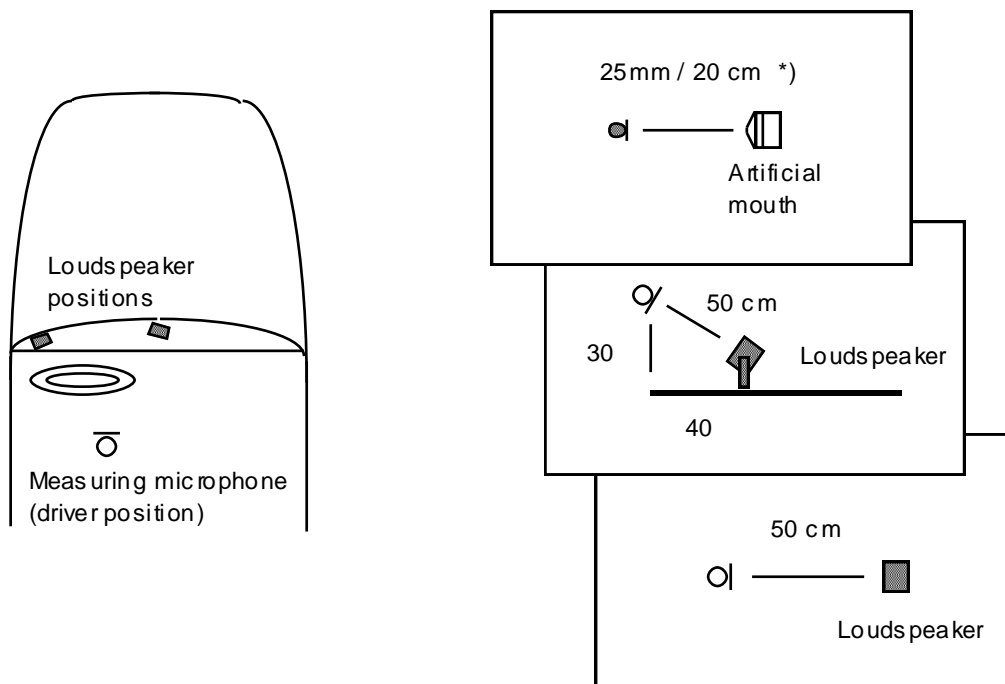


Figure 9

The sending and receiving frequency responses were measured using pink noise as test signal. The measurements were made in 1/3 octave bands.

RESULTS

The frequency response curves are shown in figure 10

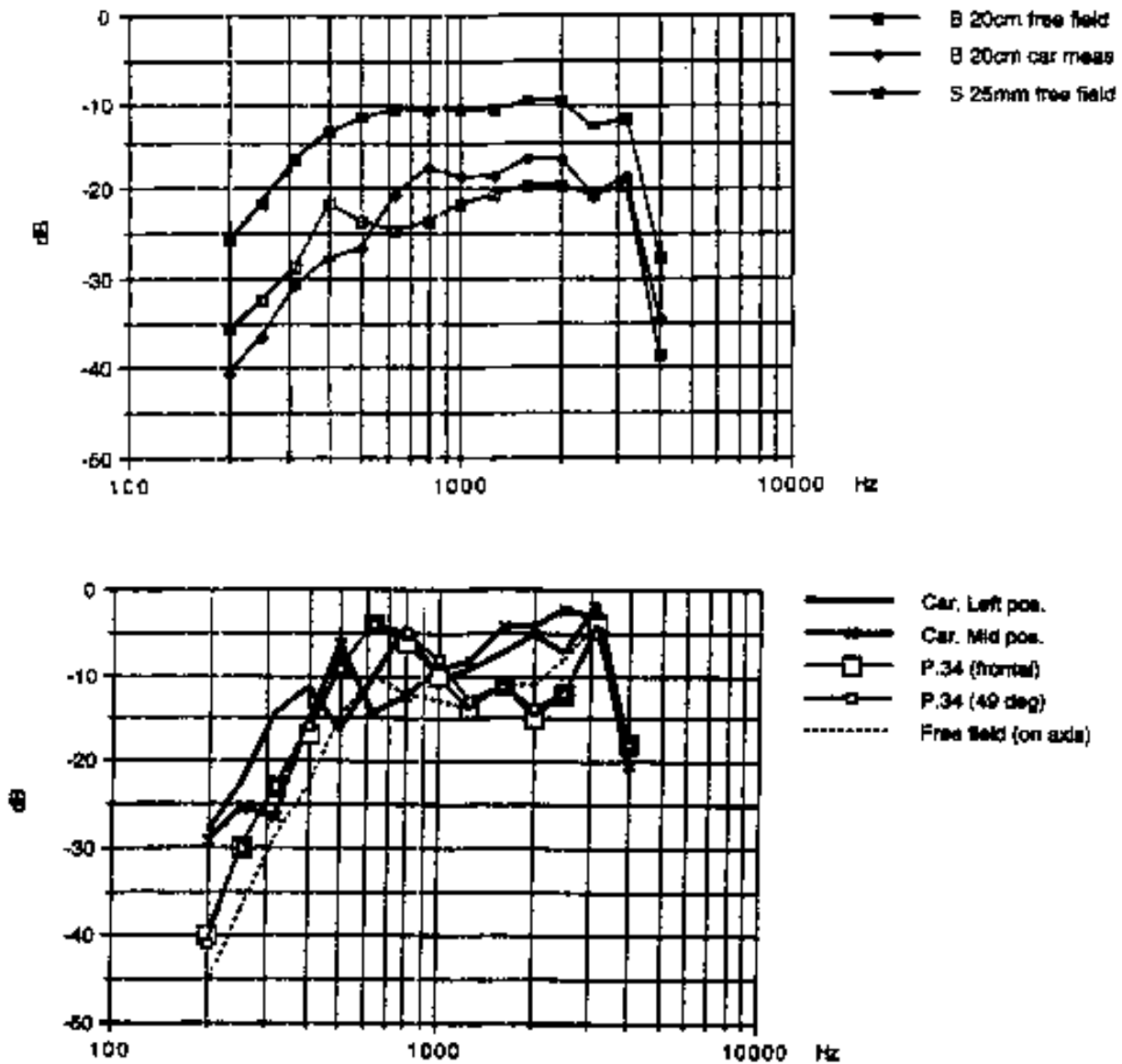


Figure 10

The first shows the frequency response at sending. The upper curve is the goose-neck microphone, the two lower ones are the chest microphone used in the car and the corresponding measurement made under anechoic conditions.

The second shows the frequency response at receiving. The two curves with fat lines are the measurements in the car with two slightly different placing of the loudspeaker (see the test set-up). The curves are equalised to correspond to a measuring distance of 50 cm. The three remaining curves are measured in free field, two of them according to P.34 with the loudspeaker placed on a flat reflecting surface, the third without the surface and with the microphone straight in front of the loudspeaker.

The corresponding loudness rating values are summarised in the tables below.

**Table 8: SLR Measurements**

Artificial mouth position	Microphone position	SLR
Driver position	Microphone 1	17,1 dB (20 cm)
Free field conditions	Microphone 1	18,0 dB (20 cm)
	Microphone 2	7,1 dB (25 mm)

**Table 9: RLR Measurements**

Measurement position	Loudspeaker position	RLR (max. volume)
Driver position	left pos	1,1 dB (corr. to 50 cm)
	mid pos	1,8 dB (corr. to 50 cm)
Ref ITU-T Recommendation. P340	(frontal)	3,3 dB
	(49 degrees)	2,8 dB
Free field conditions		6,2 dB

**Conclusion**

The frequency response at sending measured with the chest microphone hanging in free air in the car at a distance of 20 cm in front of the artificial mouth is not too far away from the corresponding curve measured under free field conditions. However, the sensitivity is somewhat higher for the chest microphone in the middle frequency region, whilst it is somewhat lower for low frequencies.

The corresponding SLR values differ by about 1 dB.

Concerning the receiving direction the sensitivity is a little higher at mid-frequencies for the loudspeaker mounted in the car when compared to the curve measured according to P.34. At lower frequencies there is no clear difference between the two cases.

The higher sensitivity at mid-frequencies is reflected in a lower RLR value for the real measurement compared to the P.340 measurement. The difference is 1-2 dB.

**4.2 Delay in handsfree terminals implemented with signal processing techniques**

**4.2.1 Signal processing techniques for acoustic echo cancelling and noise reduction**

For handsfree mobile applications, where the acoustic echo is mixed with high background noise, it is desirable to have a combination of acoustic echo cancellation and noise reduction (see subclause B.3 reference [1]). These signal processing techniques may introduce delay. Examples are given in the following subclauses.

**4.2.2 Examples of delays due to signal processing techniques**

**4.2.2.1 Data produced by Matra Communication**

In order to eliminate correctly the acoustic echoes perceived by the far-end listener due to the coupling between the handsfree loudspeaker and microphone placed in a car cockpit, the Acoustic Echo Cancellation (AEC) system needs to have efficient performances in terms of initial convergence, tracking in path variation situations, in adverse noisy and double-talk situations. Moreover the mobile network adds an important delay that will stress any bad or insufficient performances of the AEC if it is not well selected.

For this purpose it is necessary that the AEC is mainly ‘full-duplex’ in any above-mentioned conversation situations. However existing component technology implies to define a minimal additional delay allowed for echo full-duplex processing and also for additive speech enhancement processing in noise.

From simulations on real speech databases and realtime assessments, an additional parameter is defined, called *Tadd\_proc*, taking into account any additional and mandatory speech processing block (including noise bad effects compensation) where *Tadd\_proc* must be decomposed as follows:

Delay for signal block size or sub-band decomposition:	16 to 24 ms
Delay for noise reduction	12 to 16 ms
Addition delay for computation	8 to 16 ms
<i>Tadd_proc</i>	36 to 56 ms

From this global *Tadd\_proc* for speech processing block (possibly including noise reduction processing) it is also possible, and relevant in terms of extra-delay economy, to decompose the minimal requirement in three minimum values according to the kind of communication (Handset only or Handsfree). This intends to minimise the extra delays according to the kind of communications.

These new decomposition leads to three kinds of *Tadd\_proc* delays providing minimal values for acceptable echo or/and noise reduction and defined as follows:

- ***Tadd\_proc\_AEC*** for Echo Cancellation when using Handsfree MS,  
Recommended minimum *Tadd\_proc\_AEC*: 28 (to 40) ms
- ***Tadd\_proc\_NR*** when using Noise Reduction (NR) and a coupling reduction processing for use of handset MS,  
Recommended minimum *Tadd\_proc\_NR*: 20 (to 32) ms
- ***Tadd\_proc\_HF*** for Hands Free when using AEC and NR when using Handsfree MS and if it is desired to add NR for listening comfort,  
Recommended minimum *Tadd\_proc\_HF*: 36 (to 56) ms

It can be noticed that *Tadd\_proc\_HF* value is less than the sum of *Tadd\_proc\_AEC* and *Tadd\_proc\_NR* as the global processing will generally optimise the computational complexity for the association of two separate AEC and NR stages and consequently their global delay (*Tadd\_proc\_HF*).

It is clear that these extra delays *Tadd\_proc\_X* could be strongly reduced, but the extra delay values defined above correspond to signal processing responding both to realistic computational constraints and to efficient performance assessed in actual conditions of GSM communications.

To achieve a sensible extra processing delay reduction, by keeping the « full duplex » property, exponential-type increasing number of extra operations would be required: this is the case of RLS-based adaptive echo canceller working at normal full band and at sample.

Another option to have a low extra delay is to use 'Gain-switching'-based systems deploying low computational complexity. But the great disadvantages of such systems, in noisy and delayed transmissions contexts, are disastrous echo suppression and undesired switching of the useful speech to be transmitted. With such implementations double talk operation is not possible.

#### 4.2.2.2 Data produced by Ericsson

For reasons on traffic safety, it is worthwhile to use a handsfree equipment while driving a car and using a mobile telephone. The use of handsfree equipment is today legislated in some countries, and legislation is on its way in many other countries.

To obtain a good quality of conversation, the performance in this handsfree equipment should be as near full duplex as possible. The inherent long delay in today's GSM systems, max. 143.9 ms, makes the echo problem in handsfree situations much worse than in lower delay systems (e.g. ETACS, NMT, AMPS). The signal processing needed for these high quality full duplex handsfree solutions is extremely demanding. It requires a complexity which often exceeds that of today's speech and channel coding algorithms.

Block processing has proven to be a successful way, in a consumer oriented digital signal processor (DSP), to cope with the high complexity.

To fully exploit the benefits of block processing, it is important that the block length is sufficient:

- to provide enough data for statistically good estimates of the properties of noisy speech;
- to efficiently handle reverberation times in normal car cabins;
- to provide adequate resolution in the frequency domain.

It is also desirable, for reasons of efficiency in frequency domain processing, that the block length is a power of two.

A block length of 256 samples, which corresponds to a 32 ms block at 8 ksamples/s, represents a good balance in these respects.

With 32 ms blocks there will be an inherent delay of 32 ms. A reasonable figure for processing time of such blocks, using today's cost competitive DSPs, is 10 ms.

Our proposal is to allow a minimum of **32 ms + 10 ms = 42 ms** additional delay for handsfree signal processing.

## 4.3 Speech quality assessment

### 4.3.1 General - Factors affecting the speech quality of the GSM system and derivatives

#### Speech and Channel Coding Issues

One of the most fundamental parts of the GSM system is the speech codec. To reach the required spectral efficiency, the speech codec is used to compress the speech data to minimise the amount of transmitted information. To achieve this, the speech codec effectively models the vocal tract of the user as a filter and sends the filter coefficients to the decoder, in the TRAU, together with some residual excitation energy. This basic method enables the speech codec to reduce the speech data rate to 13 kbit/s for the full-rate (TCH-FS) and EFR codecs and 5.6 kbit/s for the half-rate (TCH-HS) codec.

The reduction in data rate adds distortion to the speech signal, the extreme effect of which can be heard in some of the military speech codecs, where the emphasis has been placed on intelligibility rather than speaker recognition. As a result the speech sounds 'Robotic'. The traditional measure for this distortion is the Quantization Distortion Unit or QDU. One QDU is equivalent to the amount of distortion introduced by a single transition from analogue speech to 64 kbit/s G.711 PCM and back to analogue speech. GSM Standard 06.10 states that the GSM speech codec introduces between 7 and 8 QDUs under error free (EPO) radio conditions. The QDU is an accurate measure for tandem PCM and ADPCM systems, allowing planners to determine the amount of distortion in a call routing path and ensure that it does not breach ITU-T recommended limits. ITU-T G.113 states that an international connection should not exceed 14 QDUs, this is broken down into the 5-4-5 rule. This allows the originating nation 5 QDUs, the international transit network 4 QDUs and the terminating nation 5 QDUs. Clearly a GSM to GSM call, national or international, will either just meet or breach these guidelines. Furthermore, there is some subjective evidence to suggest that the older generation, who grew up with an analogue PSTN equipped with electro-mechanical switches, will accept the 14 QDU limit. However the younger generation, who have only really known modern digitally switched fixed networks, will only accept an upper limit of 9 QDU.

Another problem with speech codecs occurs when one codec is tandemed with another of either the same or a different type. Apart from the ITU-T G.721 32 kbit/s ADPCM standard, where synchronous coding adjustment allows tandeming to occur without any further distortion to be incurred, most calls will incur more distortion when speech codecs are tandemed. This is where the QDU begins to be an inaccurate measure as very low bit rate systems do not behave in a linear way. This is because some of the speech data, necessary for the second codec to produce an accurate representation of the input speech, has already been removed by the first codec, compounding the distortion effect. Although the ITU-T have now introduced the concept at present on the validity of these measures for planning the end to end distortion of connections with low bit rate systems. In addition, the standard only includes the Impairment Factors of ITU codecs. Tests would be required to assess the TCH-FS/HS and EFR codecs.

Clearly, any improvement in speech quality is to be applauded. The recent adoption of the U.S.1. algorithm as the EFR codec has improved the speech quality of the GSM system, in error free or low error environments, to 'Wire-line' quality levels. The challenge now is to further improve on this advance to make it as invisible to the customer as possible that the telephone they are using is a cellular radio.



Most cellular network operators are currently embarked on major cell build programmes aimed to provide high levels of coverage and capacity. Even with the large numbers of cells currently foreseen there are still areas of cells where the error performance of the system is poor. This is due to a number of factors. Radio planning tools rely on clutter databases to assess the building densities in a given geographic area. These databases do not hold a picture of the actual buildings but an approximation of what is actually present. This is also true for terrain databases. This means that the actual coverage differs from the predicted coverage. Drive testing can be used to check the quality and depth of coverage in a given area but this will not correct the entire network. The buildings that populate a given area all have different radio penetration losses due to the large variety of construction methods and even decor used. This cannot be planned for by the network operator. Hence, the occurrence of EP2 (C/I 7 dB) and EP3 (C/I 4 dB) is more frequent than desired. One method of overcoming these problems is to examine the possibilities of 'Robust' coding.

There is already technology available that shows that a 'Robust' codec is possible and that can achieve 'Wire-line' quality across a broad range of operating conditions. In theory it could offer additional capacity in the network as the limit of C/I used for radio planning could be relaxed. In practice, should a 'Robust' codec be adopted, the SACCH signalling channel becomes the limit of performance. The SACCH is about 2-3 dB more tolerant to C/I than the current full-rate speech codec. An additional 2-3 dB of C/I margin would provide operators with additional fringe coverage and greater depth on in-building coverage. A major gain in radio capacity and coverage could be achieved if the SACCH performance could be improved to the same levels as a 'Robust' codec. This might provide an additional 5 dB or even 6 dB of C/I margin over the current full-rate system. The full benefits would not be realised until the majority of the operators customer base was equipped with 'Robust' mobiles, but the possibilities are worth exploring.

Discontinuous Transmission (DTX) is another aspect of speech coding that affects the speech quality of the system. The Voice Activity Detector used to detect when the customer is speaking or not, inevitably introduces some clipping which can reduce the quality. Acoustic noise also effects the DTX system. The full-rate VAD was optimised to work with in car noise where the temporal characteristics are less dynamic than say street noise. As a result the VAD can be 'false' triggered, reducing the effectiveness of DTX at reducing C/I and extending mobile battery life. Frequency Hopping is useful to reduce the effect of fading on the speech path by hopping to another channel that is not affected by the fade.

The complexity of a given speech codec has a knock on effect on the speech quality of the system. With a more complex algorithm, the speech codec can usually produce better speech quality for a given bit rate. However, to support that codec is mobile equipment, the signal processing technology has to be able to run the software in a reasonable period of time. Hence, complexity and delay are intrinsically linked. Delay, as we shall see later, can seriously degrade the quality of a connection.

### **Terminal Issues**

Many, of the key speech quality parameters are determined solely within the GSM mobile. In addition, several of these parameters are critical of the interworking between the mobile network and any interconnected networks.

The Send Loudness Rating (SLR), in the mobile to land direction, and the Receive Loudness Rating (RLR) in the land to mobile direction, determine the audio signal levels for the customers speech. The loudness ratings are calculated from the send and receive sensitivity masks or frequency responses. These are dictated by the acoustic transducers used as well as the anti-aliasing and reconstruction filters for the analogue to digital converters. One criticism levelled at GSM is that it is 'quiet'. This is difficult to understand as the SLR and RLR are in line with ITU-T long term values.

It is important that the test method for TCL as well as the requirement are carefully considered. Echo problems were reported when GSM was first introduced, even on national calls. The GSM phase 1 TCL test allowed sinusoidal test stimuli to be passed through the speech codec, as some manufacturers use aspects of the speech codec in their acoustic echo cancellation devices. Unfortunately the GSM full-rate codec causes spectral spreading on sinusoidal signals which, when measured at their discrete frequencies, have a lower power than the equivalent speech signal. This meant that mobiles were passing the test but failing in an operational environment. The phase 2 test uses the ITU-T P.50 artificial voice to address this problem.

Acoustic echo cancellation can be used as part of the TCL solution. However, to produce a stable acoustic echo canceller requires a complex algorithm due to the variable nature of the echo path. A complex algorithm requires additional delay which will cause other end to end speech quality problems. A low delay algorithm will not be as resilient but can be used as one part of the TCL solution.

The interaction of acoustic noise on other system components can have a damaging effect on their ability to meet the needs of the customers. A particular example of this is the current half-rate codec noise problems where an acoustic noise rejection mask has been demonstrated to greatly improve the performance but has been rejected as it places

stringent design restrictions on mobile terminals. An alternative algorithmic approach has been proposed but this will take as long, if not longer to implement, will require more delay and will still affect the speech quality. Noise coupling can have a serious affect of the public perception of the service offered.

Handsfree terminals pose a particular problem to speech quality. The traditional vehicle mounted handsfree environment is particularly hostile. The acoustic volume is surrounded by a reflective material with high ambient noise levels. To overcome this the handsfree system has to use some algorithmic solutions but these introduce additional delay which causes problems with conversation quality. The recent adoption (ETS 300 903) of an additional delay for handsfree processing has increased the GSM one way delay budget by 40 % and is highly likely to cause problems with end to end quality. In addition a TCL limit (lower than 46 dB) has been adopted for handsfree. The argument for reducing the TCL is that the noise will mask any echo. However, the GSM handsfree car phone is installed in an environment where several people can use it. The system is usually installed for the benefit of the driver but a rear seat passenger may wish to use it, or the passengers may all wish to contribute to a call, requiring the volume to be increased so that they can all hear the system.

To avoid future problems in the operational environment it would be better to further develop the test methods to use a Head And Torso Simulator (HATS) which more accurately represents the human that will use the terminal. In addition, carrying out a more system oriented set of tests using an artificial speech stimuli, such as the ITU-T P.50 algorithm, including the codec would be more appropriate.

### **Network Issues**

The Mobile Switching Centre (MSC) incorporates an echo canceller adhering to ITU-T G.165 with at least a 60 ms echo path window. This is because the fixed network does not have echo control in the national network and the delay of the GSM system necessitates some form of echo control. It is important that the interaction of the MSC echo canceller with other network echo cancellers is understood. In an international connection between Europe and USA there will be an echo canceller in the home countries International Switching Centre (ISC) looking at the GSM network. This canceller will normally have a 64 ms echo path window but will be trying to cancel echoes over a 190 ms echo path. In addition the non-linear GSM speech codec renders the canceller ineffective. For this reason the ISC echo canceller should be switched out of the connection. The far end ISC will have an echo canceller looking at the far end customer. This canceller is connected in tandem with the MSC echo canceller. Tests carried out by BTL have shown that echo cancellers connected in tandem can reduce the amount of echo control by between 3 dB and 6 dB dependent on whether the centre clippers are active or not. For this reason the MSC echo canceller should be switched out. The mechanism for doing this is the CCITT n° 7 signalling system. By acting upon the information on the echo flag in the Initial Address Message (IAM) and the Final Address Message (FAM), the cancellers can be correctly controlled. In addition when choosing an echo canceller it important to realise that most cancellers have been built to meet the Blue Book G.165 which uses white noise to assess the cancellers performance. Unfortunately their performance can be reduced when operating with speech.

The fixed links used to connect the BTS to the BSC and on to the MSC via the TRAU can affect the speech quality of the system. Fixed link errors can occur that can cause errors on the Abis speech frames which do not have any error correction on them. Ultimately, this could cause a bad frame to be seen as good by the TRAU. It has been suggested that TRAU bypass should be implemented to eliminate the effects of tandemed codecs. It must be remembered that this will only be true for connections between mobiles on the same coding scheme, e.g. full-rate. It should also be remembered that the same speech frame will be subjected to two radio paths. Hence, a mobile in an EP1 radio environment connected to another mobile in an EP1 environment could create an EP3 call.

Codec tandeming is also a problem for non-real time communications. Voice messaging systems also use low bit rate speech coding to store the message. This adds another type of codec and more distortion to the connection. Call forwarding can also have an effect on the speech quality. Calls from a mobile to a fixed phone which has been diverted to a mobile will not be as good quality as a call which goes directly between the two mobiles as the echo control and delay will not be optimal.

When developing interconnect agreements it is important to remember the routing of calls through the interconnected network. International calls, in particular, can be routed through Digital Circuit Multiplication Equipment (DCME) and Digital Speech Interpolation (DSI) systems. These add delay, distortion and clipping which may be unacceptable when combined with the GSM system. The international network also makes use of geostationary satellite routes which add 260 ms one way delay. GSM operators may wish to have their calls routed via cable connections as a preference.

### **Interconnect Issues**

When developing and implementing a GSM mobile network it is important that the effects of the given design on the end-to-end speech quality are understood such that a design that offers acceptable quality across the full range of call scenarios, is implemented. This means that a balance must be found between the various engineering and commercial pressures that exist when designing the system. The difficulties of apportioning the various transmission parameters becomes increasingly difficult in a multi-operator environment. In many countries there is still only one fixed network operator with one or two cellular operators and a number of Private Branch Networks (PBNs). With this scenario it is relatively easy to apportion delay, loudness ratings and the other transmission quality parameters. However, as telecommunications liberalisation progresses most countries will develop a multi-operator environment similar to that in the United Kingdom where there are over 100 licensed operators, offering local, trunk and radio networks and services.

Originally the United Kingdom relied on individual interconnect agreements between operators to agree the apportionment of transmission parameters across the network boundaries, guided by the Network Code Of Practice (NCOP). As the number of operators grew this technique became difficult to manage and the industry is now moving towards standard contracts with the transmission apportionment following the Network Performance Design Standards (NPDS) document produced for the Public Network Operators Interest Group (PNO-IG), a subgroup of the Department of Trade and Industry's (DTI) Network Interfaces Coordination Committee (NICC).

The NPDS document uses a network model to apportion transmission parameters. End-to-end limits are taken from ITU-T Recommendations and individual system standards, such as the GSM Recommendations. It is assumed that there are local loop operators, Radio Local Loop (RLL) operators, such as Ionica, cellular operators, such as Cellnet, and trunk network operators, such as BT and Energis.

One of the major principles of the document is that it is the responsibility of the operator that bills the customer, to deal with any complaints. This means that if a fixed customer calls a mobile customer, hears echo and subsequently complains to the fixed operator, the fixed operator has the responsibility to address the problem with the mobile operator. Unfortunately, in this example, the echo is usually due to the mobile customer's handset. The only safeguard that the mobile network operator has that the handset will not cause echo is that it has passed a type approval test, which we have already seen, can be inadequate and are under pressure to be removed. Current type approval test methods for digital cellular handsets do not fully represent their operational use and hence it is possible for handsets to pass the test but still cause echo whilst in use. The mobile operator has little or no control over the handsets connected to his network. The echo problems, if generated by a sufficient number of handsets and hence affect a large number of calls, can lead to litigation between the fixed and mobile operator. It is important that the type approval tests continue to be carried out and that they reflect the operational use of the handset.

### 4.3.2 Main Assessment Criteria for Handsfree processing used in GSM mobile environment

From Field tests performed with handsfree GSM mobile phone in car and office environment it appeared that the main assessment criteria relevant to assess in a realistic way various handsfree processing solutions are the following (classified into two main categories):

- \* Parameters affecting the Speech transmission quality:
  - *clarity* for users at each side of the communication
  - *distortion* on speech to be transmitted
  - *Intelligibility* of transmitted speech
  - *Double-talk behaviour* when users are talking simultaneously
  - *clipping effect*: on words end-points, hashing of words, generally characterising Echo suppresser techniques
  - *fluctuation* of the received voice at each side of the communication
  
- \* Annoying artefacts due to the Handsfree processing:
  - residual *echo* level and nature and stability (e.g. convergence/tracking)
  - *noise* level

- *noise* contrast effects

The Handsfree processing stage can include an association of an Echo Canceller (generally equalisation adaptive filters), a Gain switching system and possibly a noise reduction module.

### 4.3.3 Evaluation Methodology for Full-Duplex Acoustic Echo Controllers developed within the FREETEL-Esprit project

These data are derived from two publications referred as in reference [4] of subclause B.5 and reference [5] of subclause B.5.

For assessing a Handsfree function of a mobile/fixed telephone a clear and reliable methodology of evaluation based on the speech quality objective criteria imposed on the Acoustic Echo Controllers (AEC) is needed in order to compare and assess several candidates of AEC. Hence an evaluation methodology, preferably designed in an objective way for reliability and costs reasons, must also include the definition of the test signals that may be speech or/and synthetic, and after that the selection of a reduced database corpus performed with the realistic acoustic front-ends used in the relevant environments (handsfree mobile phones in cars, offices...).

An objective evaluation methodology for Acoustic Echo Cancellers (AEC) based on ITU-T recommendation G.167 has been developed within the FREETEL project<sup>1</sup> and allowed to evaluate/compare the AEC algorithms performances on real echo/speech signals and acoustic front-ends recordings.

However some modifications have been felt necessary in the project to adapt this objective methodology to the evaluation of a real Handsfree device with selected or given realtime integrated AEC algorithm in its context of application.

One of the main interests of this real-context objective evaluation targeted by was to offer the possibility of correlating the objective criteria measurements with partial subjective tests performed by a speaker/listener using the handsfree device in the same time.

In order to validate the proposed objective evaluation procedure for a real handsfree device it was led to dispose of a subjective evaluation methodology to assess the speech quality through the global handsfree system including:

- the Acoustic Front-End (AFE) composed of the Handsfree terminal microphone(s) and loudspeaker(s);
- the AEC algorithm with its, so-called, "glue" (corresponding to Double-Talk Detector (DTD), Voice Activity Detector (VAD));
- the GSM network, which was in this case the Full-rate codec.

Works on show, after first evaluations of AEC approaches developed in the project, some lacks of reliability on the objective criteria extracted from ITU-T recommendations for assessing handsfree functions of telephones, and, namely for GSM mobiles, adapted objective-subjective procedure based on the combination of a set of objective criteria measurements with limited subjective tests orientated towards the validation of the objective parameters is proposed.

This evaluation methodology elaborated in FREETEL project is presented in two main parts:

- Objective Evaluation procedure based on the parameters of ITU-T Recommendation G.167, including comments on the relevance of each parameter by the evaluation of Full-duplex AEC algorithms on real-contexts databases. This is described in next subclause 4.3.3.1.
- Adapted Objective Evaluation procedure based on the previous Objective Evaluation to be applied for the evaluation of any real-time echo canceller algorithm integrated in a real handsfree acoustic front-end device. This step of the Evaluation would be useful for the prototyping of a realtime echo canceller in a given handsfree acoustic front-end (AFE). This is described in next subclause 4.3.3.2.

---

<sup>1</sup>FREETEL : enhancement of handsFREE TELEcommunications, project n°6166 funded by the ESPRIT III programme.

### 4.3.3.1 Objective Evaluation procedure

An Objective Evaluation procedure is worthwhile if it complies with recommendations of telecommunications standards. A first and complete issue of ITU-T recommendation G.167 specifically applying Acoustic Echo Controllers (AEC) specifies a list of parameters to take into account for the working of an AEC placed in a given handsfree terminal for each kind of telecommunication networks, in particular for mobile radio systems. Recommended values for these parameters are generally provisional and are regularly amended according to the progress of the techniques and the real limits that would be given as feedback by handsfree telephones users: indeed it does not exist a clear specification of these limits, as also stressed in references [2], [3] of subclause B.5, however the recommended values provided by ITU-T Recommendation G.167 are precious and helpful for the AEC evaluation as they constitute the first and alone specifications with simple methodologies of tests provided in this context.

#### 4.3.3.1.1 Objective Evaluation methodology of AEC devices

The Evaluation Methodology of AECs is based on a list of AEC performances parameters defined by ITU-T Recommendation G.167. This list, including 12 AEC performance parameters that acoustic echo controllers have to comply when placed in a handsfree terminal, can be classified into 3 main classes:

- Terminal coupling loss parameters
- Time adaptivity parameters

Attenuation of speech in double-talk situations

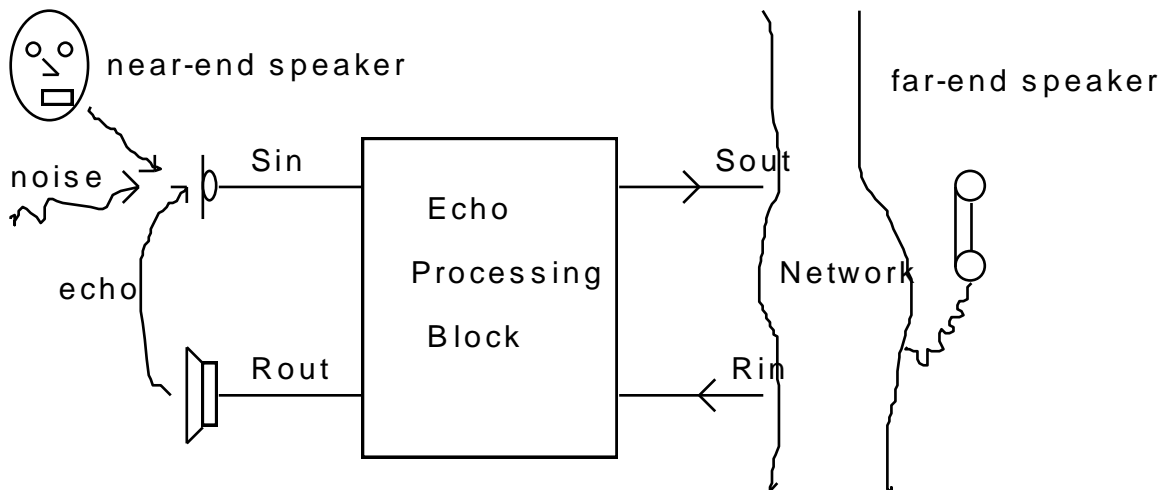


Figure 11: Handsfree audio terminal

#### a) Terminal Coupling Loss

Denoted TCL (or TCL<sub>w</sub>), is the overall attenuation of the echo resulting from the acoustic coupling of the terminal combined with the effect of the echo canceller. In fact it corresponds to the following ratio in dB,

$$TCL = 10 \log \frac{En(Rin)}{En(Sout)} \tag{1}$$

where En(.) defines the energy of respectively the Received input signal (Rin) and the Sent output signal (Sout) at ends of the audio terminal. In fact TCL (or TCL<sub>w</sub>) corresponds to the sum of two important quantities, the coupling loss between the loudspeaker and the microphone, denoted CL, and the Echo Return Loss Enhancement, denoted ERLE, which measures the intrinsic efficiency of the AEC algorithm.

The coupling loss (CL) has the following expression:

$$CL = 10 \log \frac{En(Rout)}{En(Sin)} \quad (2)$$

ERLE is the difference between the raw echo and the reduced echo got as output of the AEC processing whose expression is:

$$ERLE = 10 \log \frac{En(Sin)}{En(Sout)} \quad (3)$$

with,

$$TCL = CL + ERLE \quad (4)$$

if by assuming that  $Rin = Rout$ , i.e. no loss is present in the Receive way (case of an AEC).

In the case of loss in the Receive way of the handsfree system - use of an echo suppresser (realised by gain switching) or more generally existence of attenuation/gain in the receive part - we must take into account this third quantity in the parameter TCL, the Receive Loss, denoted RL,

$$RL = 10 \log \frac{En(Rin)}{En(Rout)} \quad (5)$$

which gives in the general case,

$$TCL = RL + CL + ERLE \quad (6)$$

In the Evaluation Methodology it is taken into account more specifically the ERLE quantity (relative to the intrinsic performance of the AEC stage for a given acoustic front-end) as the TCL can be deduced from it by adding the RL and CL parameters that are generally considered as constant (RLcst, CLst) in the case of digital Handsfree functions. However for Echo Suppressers (ES) the RL quantity is not exactly constant but it can be noted that it is composed of a symmetric value of the ERLE plus a constant RL value. This can be summarised as follows:

$$\text{- for AECs: } TCL = ERLE + (RLcst + CLcst) \quad (7)$$

$$\text{- for ESs: } TCL = 2.ERLE + (RLcst + CLcst) \quad (8.1)$$

where

$$RL = ERLE + RLcst \quad (8.2)$$

From this, three main sub-quantities characterising the AEC in three specific situations as simple talk (st), double talk (dt) and echo path variation (pv) are defined:

**a.1. ERLE<sub>st</sub>/TCL<sub>wst</sub>** which measures the terminal coupling loss in Simple Talk situation, i.e. the far-end subscriber is talking without near-end signal simulating the near-end speaker or the Handsfree terminal user.

It is recognized that it is necessary to reach a minimum value of 46 dB in case of important network delays (e.g. GSM) and, on the other hand, it is really difficult to obtain such a high value without degrading the transmission quality, with the existing Echo Suppressers based on gain switching technology, in particular in transient situations as "dt" and gain switching phases in general. Indeed this deficiency of Echo Suppressers (Gain Switching) is moreover stressed by the important Networks delays.

Now it is clear that such a limit for TCL<sub>wst</sub> is still considered as difficult and consequently in a full-duplex way as AEC algorithms cannot presently reach this value without unrealistic big computational load, which, then, precludes any integration of such complex algorithms on existing component technology fitting to manufacturing constraints of GSM terminals. It should seem more realistic to reduce this limit for handsfree terminals to take into account the real efficiency of AEC techniques that can be deployed with existing and cost-effective Signal Processing Components Technology. This was the case in the actual GSM Standard ETS 300 903. These figures represent a fair trade-off between maintained 'Full-duplex' properties (requested by users) and an acceptable level of residual echoes reduction.

This parameter after successive evaluations was found as a reliable objective criteria pending that test signals are close to real speech signals.

**a.2. ERLEdt/TCLwdt** corresponding to the previous case but with near-end speaker activity

It is recognized that a TCL (and TCLw) limit in double talk situation could be lower than in single talk (Preliminary values are available in I-ETS 300 245-3).

Evaluation Methodology implemented the proposed scheme and obtained results unable to separate correctly different AEC candidates. Other approaches to compute the ERLEdt were proposed in reference [4] of subclause B.5, but no clear conclusions were drawn, except that this parameter was not easy to represent. Hence it has been provisionally concluded that only subjective measurements could be performed.

**a.3. ERLEpv/TCLwpv** measuring the echo loss during echo path variation: the AEC is let to converge to value ERLEst (or TCLwst) in the "st situation", then an echo path variation is applied during 5s, the ERLEpv is the ERLE measurement at the end of this time.

TCLwpv should be *greater than 10 dB*.

**a.4. ERLEstn/ERLEpvn/ERLEdtn (TCLstn/TCLpvn/TCLdtn)** measuring the echo loss during simple talk, echo path variation or Double-talk in noisy environments (i.e. when noise is added to the echo, or to the echo plus the near-end speech). The ERLE<sub>xxn</sub>/TCL<sub>xxn</sub> evaluation procedures in noise have been kept similar to a.1, a.2, and a.3 procedures, but values here need to be defined. Moreover, depending on its spectral shape, noise may present different masking effects on the echo. So values must be defined for a given kind of noise to be selected among: white noise, low-frequency noise (car noise), medium frequencies noise (town & cars traffic) or speech-like noise (babble, cocktail party).

In order to get reliable values of TCL in noise, better is to use noise with stationary spectrum like car noise or (filtered-) white noise.

Even if values are not defined, the ERLE<sub>xxn</sub>/TCL<sub>xxn</sub> evaluation in noise is recommended and allows to check and compare the robustness properties of different AEC techniques.

## **b. Time adaptivity Parameters**

Time adaptivity parameters are intended to evaluate the performances of an AEC in different situations of communications. Five parameters have been identified in ITU-T Recommendation G.167 corresponding to four precise communication modes:

**b.1. Tic** (initial convergence time) characterises the convergence behaviour of the AEC at the beginning of a communication assuming that the near-speaker does not talk or has already talked a first time. In ITU-T recommendation G.167 the echo attenuation in simple talk measured by the TCLst should be *at least 20 dB after Tic=1s*.

This parameter has been integrated in the Evaluation Methodology and provided a quite reliable measurement pending that test signals are energetic enough on a sufficient time period and close to real speech signals (see Test Signals Database in subclause 4.3.3.1.2).

**b.2. Trdt** (recovery time after double-talk) is the time necessary to recover a given echo attenuation after a double-talk event.

The provisional limit in ITU-T Recommendation G.167 is *Trdt=1s for a ERLEst of at least 20 dB*.

In practice the double-talk is generated during 2s after the AEC has reached TCLwst, then cut off and a timer is started to measure the re-convergence of the EC in simple-talk again (after double-talk event). This parameter was also integrated in the Evaluation Methodology but no conclusive results can be drawn.

**b.3. Trpv** (time necessary to recover a given attenuation of the echo after an echo path time variation).

*Trpv should have the same limit as Trdt.*

The means to produce an echo path variation is still under study by ITU-T, so the "gain varying" technique applied on the echo present in the microphone channel was provisionally considered as a quite representative of an echo path variation in the Evaluation Methodology.

**b.4. Tonst/Tondt**(break-in times in simple and double talk situations) are intended to measure the decreasing speed of transmitted speech attenuation. These parameters, computed for both Receive and Send sides, are more relevant to the Echo Suppressers case (half-duplex gain switching techniques).

From Tonst, when attenuation becomes lower than 3 dB, Tonst should be < 20 ms. On the other hand Tondt should be < 20 ms when attenuation becomes lower than 6 dB.

Hence in the Evaluation Methodology these parameters have not been retained as AEC techniques are full-duplex and no switching gain are applied on each channel. Nevertheless these quantities should be added if a full-duplex AEC (Adaptive filtering) is associated with an Echo Suppressor.

**b.5. Tic\_n/Trdt\_n/Trpv\_n** measuring the Time adaptivity or reactivity of the AEC techniques during simple talk, Double-talk or echo path variation in noisy environments (when noise is added to the echo, or to the echo plus the near-end speech). The corresponding evaluation procedures in noise have been kept similar to b.1, b.2, and b.3 procedures, but values here need to be defined and equivalent remarks on noise shapes to TCLxxn in a.4 are also applying.

#### c) Speech Attenuation/Distortion in Double Talk mode

These parameters are intended to measure the attenuation and distortion observed onto received and sent speech when double talk situations occur. These parameters are respectively denoted Ardt/Asdt and Drdt/Dsdt for receive/send speech attenuation and distortion.

These parameters have not been used in the Evaluation Methodology as their interpretation has been found problematic for the case of AEC based on digital filtering (see subclause B.5 reference [4]).

**c.1. Ardt/Asdt** correspond to the speech attenuation (received/sent speech) during double talk. A similar evaluation procedure as for TCLwdt can be used. Asdt/Ardt are obtained by comparing the speech just before adding 2s of double talk and the attenuated speech just after removing the double talk signal.

In both sides, Asdt and Ardt should be less than 6 dB.

**c.2. Drdt/Dsdt** correspond to the distortion of received/sent speech during double talk. The evaluation procedure is similar to the TCLwdt one, the distortion is obtained by comparing transmitted speech before and after double talk in terms of a distortion measure.

Values are not specified and they are still under study.

NOTE: For AECs Ardt and Drdt are not necessary since AECs do not alter speech signal in the receive way both for simple talk and double talk situations.

#### 4.3.3.1.2 Test Signals used from the FREETEL-Esprit Database

Test Signals have been extracted from the FREETEL-Esprit database including handsfree telephone sets conditions for GSM applications in car and room (offices): handsfree acoustic part of the terminal was used for office echo recordings whereas separated microphone-loudspeaker(s) handsfree front-end was selected for echo recordings in car (Renault 25). In fact this last handsfree front-end corresponded to the commonly named "handsfree car-kit" now commercialised for most of the GSM mobile placed in car.

Several kinds of signals have been recorded. These signal sequences include at the same time the far-end signal recorded at the loudspeaker input, (i.e. Rout) and the echo signal picked up at the microphone output (i.e. Sin):

- real speech signals (1M/1F) for both English/French languages including couples of sentences and balanced energetic sentences for measuring AEC convergence/echo reduction performances such as (Tic, Trpv) and averaged (TCLst, TCLpv).
- USASI signals which correspond to long-term average of speech signals and often used for their continuity of energy. They are very useful for tuning the AECs and also measuring (TCLst, TCLpv) parameters.
- White noise sequences for the acoustic echo impulse response identification and for the analysis of transfer frequency function between the loudspeaker and the microphone.



- CSS signals (Composite Source Signals) which are composed of high-, low-correlated signals and silence periods aiming at simulating speech signals. The use of real speech (couples of sentences & energetic sentences) or constant synthetic signals as USASI were preferred in this Evaluation Methodology.

NOTE: This part of the FREETEL database is not presently available but a procedure to make it "public" is under way.

#### 4.3.3.2 Adapted objective evaluation procedure to a GSM Handsfree mobile

##### 4.3.3.2.1 Motivation of an adapted evaluation procedure

In the context of the GSM mobile the objective methodology of evaluation presented in subclause 4.3.3.1 and in details in reference document [4] of subclause B.5, had to be focused onto the existing Handsfree device real-time platforms aiming at prototyping Industrial handsfree echo cancellation devices. The DSP board (EVM or commercial board), implementing the pre-selected AEC algorithms (after simulation), is connected through interfaces to the acoustic front-ends (the "Handsfree Adapter") and the PC-AT driving the application. The Handsfree Adapter allows to connect different acoustic front-ends such as a mobile handset or a 'handsfree car-kit' and permits to configurate the global system Acoustic front-end/PC-DSP in two possible ways of communications: internal mode or real GSM mode.

On the other hand the Evaluation Methodology of AEC devices has been materialized into an Evaluation Tool (ET) implemented in C-language (see subclause B.5 reference [4]) based on each of the ITU-T G.167 parameters presented in subclause 4.3.3.1.1. However some modifications in the ET were felt necessary namely by only keeping objective parameters that were found reliable and relevant for AEC during preliminary evaluations on real acoustic front-ends signals for pre-selection of AEC algorithms. The separameters ??? are::

- echo reduction measures in Simple talk ('st') and Path Variation ('pv') modes: ERLEst/TCLst, ERLEpv/TCLpv, Tic, Trpv;
- the similar parameters but computed in noisy conditions: ERLExxn/TCLxxn, Txx\_n.

The procedure is based on the use of the far-end terminal either in the handsfree or in the handset modes in order to take into account the possible distortion effects onto the far-end speech signal Rin going across the different communication stages inside the mobile terminal (A/D-D/A converting, speaker gain, speech encoding/decoding...) till the loudspeaker input Rout of the near-end terminal still in the handsfree mode. This procedure, presented in the following subclause 4.3.3.2.2, is materialised by recordings of useful signals for the echoes characterisation and AEC processing evaluation: the objective is to perform these recordings by using the global system PC-DSP-Handsfree Adapter above-described.

On the other side the GSM network effect, which is essentially

- the 90 ms delay making a double-delay shift onto the acoustical echo versus to the original signal,
- the low-level noise signals, generated by the GSM network and that may perturbate the AEC filter,

had to be included and, this, by performing a real GSM communication between the near-end and the far-end parts: this facility was offered by the Handsfree adapter that provided an interface with a GSM base.

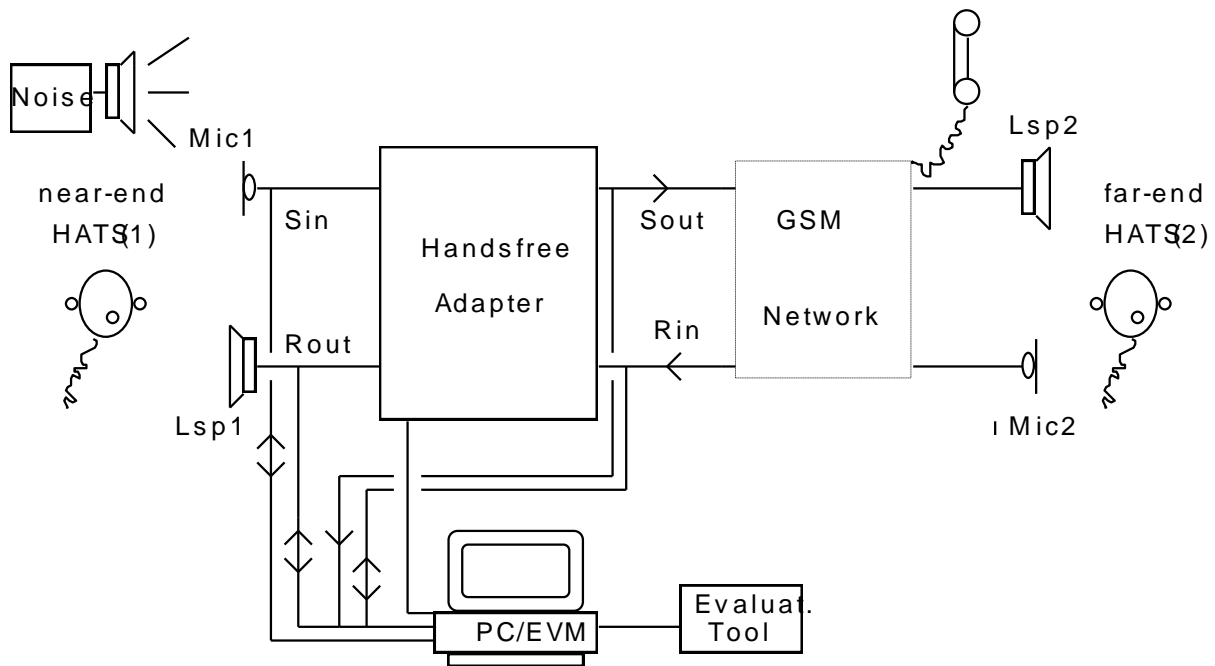
The following subclauses present this Adapted Evaluation procedure to be used to prototype any echo cancellation stage for a given acoustic front-end such as a handsfree car-kit (microphone and loudspeaker are independent) or a « self-contained handsfree » handset (the microphone and the loudspeaker are in the same box and close to each other). Then it is derived and proposed from this procedure a scheme for assessing a whole Handsfree device for GSM terminals.

##### 4.3.3.2.2 Objective evaluation procedure for prototyping the GSM Handsfree AEC algorithms

The proposed procedure was based on the use of a Head and Torso Simulator (HATS) successively placed in front of the different kinds of terminals connected to the Handsfree Adapter. The procedure decomposes in two main steps:

- a) creating of a specific handsfree speech database within the GSM mobile platform context
- b) objective evaluation of the AEC processing by using pre-recorded speech signals resulting from step a)

The following block-diagram provides an illustration of the procedure tool.



**Figure 12: Block-diagram of the Evaluation system for a Handsfree Mobile/Fixed Terminal**

**a) Pre-recording and storage of the specific handsfree speech database**

This phase corresponding to the pre-recording of handsfree speech materials will be only done once. It is composed of three steps,

**Step 1:** Diffuse by means of HATS (2) the speech input files of the FREETEL -Esprit database (see subclause 4.3.3.1.2, loudspeaker input files) corresponding to speech/synthetic signals. In the same time these signals are recorded both at the Lsp1 loudspeaker input, denoted Rout, and at the Mic2 microphone output, denoted Rin, in a digital format and stored into the hard disk memory of the external PC driving the DSP board. Rin can be recorded either at the Handsfree adapter interface (input of the echo cancellation system), including the GSM network, or at the far-end Mic2 output. Recordings can be synchronised at the far-end with the near-end.

**Step 2:** Newly stored signal Rout is then transformed according to a linear rule of gain varying for simulating the echo path variation. New transformed signal will be denoted Rout\_pv. All these operations/computations are supported by the PC.

**Step 3:** Signal Rout (or Rout\_pv), stored in the PC memory, is diffused by the PC through the loudspeaker Lsp1, the Sin signal at the microphone Mic1 output is recorded and stored in the PC memory. For this step of the database recording two kinds of situation are foreseen:

- in simple talk situation with the near-end HATS (1) in silent state,
- in double talk with HATS (1) diffusing in the same time a near-end speech.
- both situations with ambient noise generation extracted from a library of pre-defined noise files

**b. Objective evaluation of the AEC processing on the Handsfree Database**

After the specific database recording task realised once in a. the objective evaluation of AEC algorithms implemented into the DSP can be started every time one desires. During the AEC processing, the following operations will be done in the same time:

- the pre-stored Sin signal (echo + possible near-end speech produced by HATS(1)) is being diffused
- the pre-stored Rin signal is being diffused

- the output signal from the AEC denoted Sout in the figure will be provisionally stored into the PC memory. We recall that the output signal Sout will represent in this case the residual echo signal + possible transformed "dt" signal and environment noise.

After that the Evaluation Tool (ET) will provide by using all necessary stored signals the AEC performance parameters previously defined according to 4.3.3.1.1 and reference [4] of subclause B.5 and the deviation to the ITU-T recommended values. In fact the ET software will be preferably placed in the PC and will give these results under a synthetic table format every time one desires to test the AEC algorithms integrated inside the DSP board. Then these results can also be compared to the simulation results obtained with the same AEC methods implemented in C-unix. Nevertheless only the set of parameters retained in 4.3.3.2.1 are considered with reliance.

**This objective evaluation procedure will allow to check preferably the performances of the real time AEC versions with regards its simulated version**

#### 4.3.3.2.3 Proposed Objective-Subjective Evaluation procedure

The proposed Objective-Subjective evaluation consisted in performing listening tests by using the Handsfree system Objective ET platform presented in subclause 4.3.3.2.2. A limited number of expert listeners (e.g. 10 listeners) will give quality marks well targeted on the main handsfree effects to be analysed, previously defined in subclause 4.3.2 such as echo disturbance, bad tracking/convergence on the echo, clipping effects, transmitted ambient noise contrasts. The Handsfree system ET will compute in the same time the retained Objective parameters. This is intended to provide some feedback on the objective parameters thanks by the evaluation of subjective quality performances done by the expert listeners.

This Objective-Subjective Evaluation procedure was not fully validated but was partly used for real implementation of AEC techniques on target DSPs. This might be used also to validate firstly-retained objective parameters (in 4.3.3.2.1) and make necessary modifications/refinements on those that are not easy to interpretation/implementation (double talk situations). If this first step is considered as satisfactory then an objective evaluation based on the pre-validated objective parameters could be only done for each handsfree function to be tested.

From experiments performed with the FREETEL handsfree system in real GSM communications it has been felt mandatory to keep, for the Handsfree device evaluation, the GSM network component since it has a significant influence onto the AEC filtering stages.

#### 4.3.3.2.4 Derived Evaluation schemes for Handsfree mobile telephones

From the previous objective evaluation procedure of Handsfree AEC processing devices it is also interesting to derive the same type of procedure for Handsfree GSM terminal products.

This procedure permits to compute a set of 11 objective parameters retained in subclause 4.3.3.2.1 aiming at validating the main performances of the AEC device integrated in the Handsfree terminal.

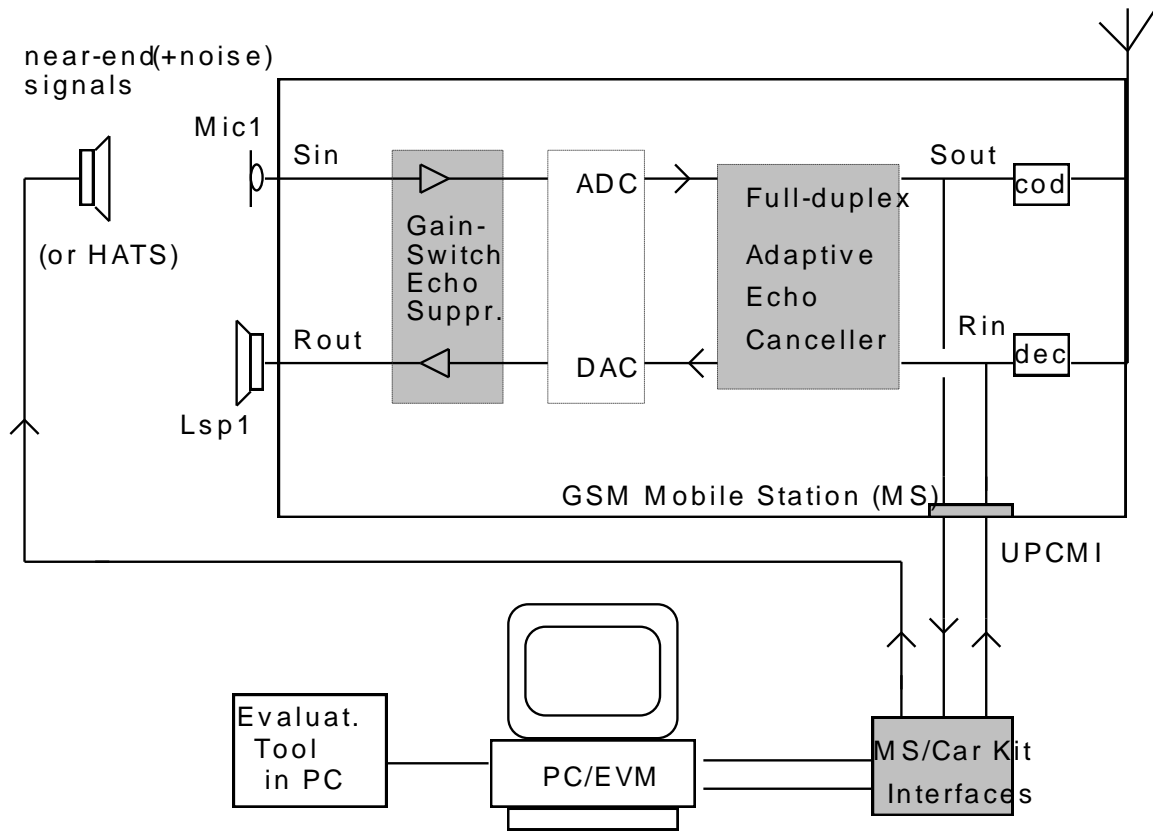
This set of 11 parameters is computed by means of the Evaluation Tool software mentioned in subclause 4.3.3.2.1 and are summarised hereafter:

- Average echo loss (TCLst) and convergence times in simple talk Tic 20 (till TCL reaches TCLst -20dB) and Tic 10 (till TCL reaches TCLst - 10 dB). Tic characterises the reactivity of the AEC when the communication starts.
- Average echo loss (TCLpv) and convergence times after echo path variation Tic\_pv\_10 / Tic\_pv\_20 (at TCLpv=10dB/20 dB) Tic\_pv measures the reactivity of AEC after "path variation" of the acoustic echo
- Performances of echo reduction in noisy environments: (TCLstn, Tic\_stn) and (TCLpvn, Tic\_pvn)

Noise is added on the acoustic echo and must present preferably a stationary spectrum.

- Behaviour assessment of a purely "Half-duplex" device (gain switching) in interactive situations between the near-end and the far-end users: the Tonst parameter measure is recommended to be added but this parameter need to be analysed.

The following schemes of evaluation can be used for measuring performances of AEC systems integrated in GSM terminals.



**Figure 13: Evaluation of the AEC function of a GSM Terminal**

Same procedure can also be used for GSM terminals equipped of a Handsfree car-kit:



time, mutilation of speech by the action of voice operated devices; which are typically degradation's for handsfree telephone. Listening only tests or one-way listening tests have been extensively used when the degradation factors to be studied affect only the listening of customers.

#### 4.3.4.1.1 Listening opinion tests

Several methods are recommended and fully described in Annex B to Annex F of ITU-T Recommendation P.800. Among these methods, the two most frequently used are the ACR (Absolute Category Rating) and the DCR (Degradation Category Rating) methods which are both category-judgements methods. One collects the subject's vote on a five-point Quality Scale (ACR), the other on a five-point Degradation Scale (DCR). For both procedures, a strict control of all experimental factors (from speech samples used for recordings to orders of presentation) through experimental design allows to obtain reliable results, which are, thus, statically validated. Naive or non-experienced subjects are used for the evaluation.

#### 4.3.4.1.2 Conversation Opinion Tests

This procedure is an interactive evaluation procedure involving two participants: a talker and a listener. They are both placed in a controlled environment and invited to hold conversations with the help of a proposed conversation task. At the end of this exchange, they are individually asked to assess the perceived quality of the communication. Various 5-point category judgements are used; two opinion scales are recommended in P.800, Annex A: the classic Quality scale and a Difficulty scale (binary response to the question: "Did you or your partner have any difficulty in talking or hearing over the connection ?")

Certain listening laboratories use more than two scales, typically a questionnaire with half a dozen items. CNET recommends a multicriteria approach where descriptive criteria are grouped together and the use of a minimum of a quality scale and of two impairment scales, namely an echo annoyance scale and a noise annoyance scale. The impairment scale is taken out the DCR method but its application to the conversation situation differs from the recommended procedure for the listening tests in that no explicit good quality reference is introduced prior to each evaluation. The experimental design followed to run this type of tests should be the  $n \times n$  graeco-latin square, suitable to take into account four variables with the same degree of freedom ( $n-1$ ): test condition, order of presentation, conversation task and subject. It is necessary to stress the point that the choice of an experimental design and of the subsequent model for variance analysis determine completely the effects which could be estimated from a statistical point of view.

In its normalised version for purpose of overall quality evaluation, non-experienced subjects are used for these tests. It is possible to conduct simplified conversation tests with different applications: diagnostic or validation of the device under test. In this specific case, professional subjects (operators of telephony, experts,...) give their opinion on several criteria; no graeco-latin square design is requested, only one conversation task is used. These simplified tests are not normalised and not statistically validated (4 or 5 collected votes only).

#### 4.3.4.1.3 Proposed test method

The proposed methodology adapted from the Ultimate Test Set is not normalised and even not yet validated. This set of tests should be considered as diagnostic tests such as those run during the development phase of a coder which are not documented in ITU-T. The procedure uses five point category-judgement scales which should be conformed to recommended existing ones: quality, listening effort (and derived talking effort) impairment scale, difficulty scale, loudness preference, etc. The fact that no experimental design, and thus, no control of experimental variables are given, along with the small number of collected votes do not allow a statistical justification of the results.

It should be noted that, within ITU-T SG 12, a new recommendation which requires similar methodological clarification, is under study in Question 14/12. The future recommendation (P.SEEC Speech Evaluation of Echo Canceller) is devoted to the evaluation of the quality performance of echo cancellers and will form a basis of work on other active signal processing devices as those implemented in handsfree terminals. A provisional list of subjective possible procedures along with their applications has been already elaborated which try to prioritize the different methodologies, e.g. Conversation Tests, Talking & Listening Tests, Third Party Listening Tests, depending of their specific goals, e.g. Overall Opinion and/or Quality, Diagnostic, Parameter Value Selection.

It is clearly said in the reports of ITU-T SG12 Question 14/12 on this matter that the unique reference will always be the "true, classic" conversation test, as the only one test and that the relationship between Conversation MOS scores and other results should be understood and firmly stated.

### 4.3.4.2 Subjective tests extracted from "Ultimate test set"

The following subclause presents a summary of Race R2072 WP2.2 contribution titled "The Ultimate Test Set".

This text defines several tests:

- Objective tests
- Subjective tests:
  - Talker test comprises tests that should reveal any degradations that will hinder the talker in speaking.
  - Listener test aims at revealing any degradations that will hinder the listener in listening.

Part 1 is described in the following (speech and silence; CVC). Part 2:see ITU-T Recommendation.

- Interactive test (interruption and interaction) is designed to assess the communication link when talker and listener are involved in a highly interactive test, where the respective roles of talker and listener are frequently swapping.
- Conversation test should reveal the subjective quality of the link when it is used in a normal conversation. (See ITU-T Recommendation P.800)

The subjective tests, except those defined in the ITU-T recommendations are described in annex A.

#### 1) Talker test

**Table 10: Talker test**

Test Type	Parameters tested
Free monologue	Noise
Counting test	Echo

The test can be performed by experimented subjects. It is recommended that at least four different talkers (two male, 2 female).

The listener can be replaced by a HATS.

#### 2) Listener test

**Table 11: Listener test**

Test type	Ambient noise	Parameters tested
Speech and silence		Echo
Speech and silence	distant ambient noise	Influence of the distant ambient noise
Logatoms (CVC)		DAV

The test can be performed by experimented subjects. It is recommended that at least four different talkers (two male, 2 female).

The talker can be replaced by a HATS placed in the standard conditions.

#### 3) Interaction and interruption tests

The test is made between two partners. It is mainly performed to test the behaviour of speech processor included in the handsfree terminal. Both subjects have active parts, they are faced to two simultaneous tasks, ex: counting and listening. This is difficult task which cannot be asked to naive subjects. Moreover, the evaluation criteria related to voice suppression, voice clipping, etc, is out of the possible interpretation of non experienced subjects.

According to the classification achieved in ITU-T SG 12 Question 14/12 , this test should remain a diagnostic tool performed by experts.



# Annex A: Subjective Tests from UTS

## TALKER TEST / SLOW COUNTING

The purpose of the test is to reveal degradations that hamper the talker in speaking. The talker is asked to slowly count out from 1 to 20 and to answer to the following questionnaire:

Q1. *Did you hear any echo ?*

Score	Echo annoyance
5	degradation is inaudible
4	degradation is audible but not annoying
3	degradation is slightly annoying
2	degradation is annoying
1	degradation is very annoying

Q2. *How would you rate the quality of the connection from the point of view of a talker?*

Score	Quality Scale
5	excellent
4	good
3	fair
2	poor
1	bad

## TALKER TEST / FREE MONOLOGUE

It is asked to have a free monologue. You are completely free in choosing whatever you find meaningful to test the quality. For example, you could whisper, make an exclamation or just talk in a normal manner. While talking, you should listen for any degradations causing difficulty in talking.

Q3. *What is your opinion on the amount of effort required to talk caused by the presence of any noise: hum, clicks, beeps or tones?*

Score	Talking effort scale
5	complete relaxation possible, no effort required
4	attention necessary, no appreciable effort required
3	moderate effort required
2	considerable effort required
1	extreme effort: talking impossible

Q4. *How would you rate the quality of the connection from the point of view of a talker?*

Score	Quality scale
5	excellent
4	good
3	fair
2	poor
1	bad

**LISTENER TEST / SPEECH AND SILENCE TEST**

You are going to hear 4 samples of speech and silence. After listening to each sample, answer to the questionnaires.

Q 5. *Did you hear any echo?*

Score				
Sample 1	Sample 2	Sample 3	Sample 4	Impairment scale:Echo annoyance
5	5	5	5	degradation is inaudible
4	4	4	4	degradation is audible but not annoying
3	3	3	3	degradation is slightly annoying
2	2	2	2	degradation is annoying
1	1	1	1	degradation is very annoying

Q 6. *How would you rate the quality of the connection from the point of view of a listener?*

Score				
Sample 1	Sample 2	Sample 3	Sample 4	Quality Scale
5	5	5	5	excellent
4	4	4	4	good
3	3	3	3	fair
2	2	2	2	poor
1	1	1	1	bad

Q 7. *What is your opinion on the amount of effort required to talk caused by the presence of any noise: hum, clicks, beeps or tones?*

Score				
Sample 1	Sample 2	Sample 3	Sample 4	Listening effort scale
5	5	5	5	complete relaxation possible, no effort required
4	4	4	4	attention necessary, no appreciable effort required
3	3	3	3	moderate effort required
2	2	2	2	considerable effort required
1	1	1	1	extreme effort

Ambient noise is generated in the distant room

Q'. *How do you perceive the distant ambient noise?*

Score				
Sample 1	Sample 2	Sample 3	Sample 4	Impairment scale: distant ambient noise annoyance
5	5	5	5	degradation is inaudible
4	4	4	4	degradation is audible but not annoying
3	3	3	3	degradation is slightly annoying
2	2	2	2	degradation is annoying
1	1	1	1	degradation is very annoying

**LISTENER TEST / CVC**

The test is designed to test the Voice Activity Detector (VAD)

*Q 9. What is your opinion on the amount of listening effort?*

<b>Score</b>	<b>Listening effort scale</b>
5	complete relaxation possible, no effort required
4	attention necessary, no appreciable effort required
3	moderate effort required
2	considerable effort required
1	extreme effort

**INTERUPTION TEST**

At the other side of the telephone connection, a talker is continuously generating speech. You are asked to interrupt the talker by uttering sounds like "uuhu", "yes", "mmm",...to detect degradations caused by the equipment.

*Q 10. How do you rate the amount of voice suppression ?*

<b>Score</b>	<b>Voice suppression perception</b>
5	no voice suppression perceived
4	voice suppression is audible but not annoying
3	voice suppression is slightly annoying
2	voice suppression is annoying
1	voice suppression is very annoying

*Q 11. How would you rate the overall communication quality ?*

<b>Note</b>	<b>Quality scale</b>
5	excellent
4	good
3	fair
2	poor
1	bad

**INTERACTION TEST/ COUNTING TEST**

The two partners are at each side of the link and are asked to take turns uttering the numbers from 1 to 20 (one uttering the even, the other the odd numbers) as quickly as possible. The counting can be repeated as often as the partners consider necessary. While counting they should listen for degradations caused by voice clipping and delay.

*Q 12. How would you rate the amount of voice clipping?*

<b>Score</b>	<b>Amount of voice clipping</b>
5	no voice clipping perceived
4	voice clipping is audible but does not influence the task
3	voice clipping slightly causes difficulty to perform the task
2	voice clipping causes considerable effort to perform the task
1	voice clipping causes extreme effort: no conversation possible

*Q 13. Do you perceive any delay?*

<b>Score</b>	<b>Amount of delay</b>
5	no delay perceived
4	delay perceived but does not influence the task
3	delay slightly causes difficulty to perform the task
2	delay causes considerable effort to perform the task
1	delay causes extreme effort: no conversation possible

*Q 14. How would you rate the overall communication quality?*

<b>Score</b>	<b>Quality scale</b>
5	excellent
4	good
3	fair
2	poor
1	bad

---

## Annex B: Bibliography

### B.1 References of TD presented in SMG2/ad hoc 03.50 meetings

- 30/95. FT/CNET/ Monfort: Analysis of information available on handsfree in mobile (vehicle) environment
- 9/94. Telia/ Karlsson: Measurement on a GSM handsfree telephone)
- 58/95. BT/ Goetz: Handsfree test environment
- 42/95 FT/CNET/ Monfort: Test to be applied to ETS 300 540
- 49/95. BT/ Goetz: Advisory text for installation of handsfree MS in a vehicle environment
- 60/95. Head Acoustics/ Gierlich: Cost estimation for a vehicle simulator for handsfree testing
- 12/95. Matra Communications/ J. Boudy: Additional speech processing delay for the use of Handsfree
- 23/95. Matra Communications/ J. Boudy: Delay figures for HF processing
- 35/95. Matra Communications/ J. Boudy: Additional delay for HF processing
- 77/95. Ericsson: Extra delay for handsfree signal processing
- 41/95 Matra Communications/ J. Boudy: Evaluation methodology for full duplex acoustic echo controllers developed within Freetel Project
- 11/94. BT/I.Goetz: Assessing GSM speech quality

---

### B.2 References from subclause 4.1

- [1] **Background acoustic noise reduction in mobile telephony.** R.A. Goubran, H.M. Hafez. 36<sup>th</sup> IEEE Vehicle Technology Conference.
- [2] **Speech enhancement for mobile telephony.** M.M. Goulding, J. Bird. IEEE Transactions on vehicular technology, Vol 39, N°4, November 1990. p 316-326.
- [3] **Contribution à l'amélioration des performances d'un radiotelephone mains-libres à commande vocale.** C. Baillargeat. Thésis (18 April 1991).
- [4] **Acoustic noise analysis and speech enhancement for mobile radio applications.** N. Dal Degan, C. Prati. Signal Processing 15, N°1, July 1988, p43-56.
- [5] **On the influence of front end processing schemes on the GSM codec behaviour in the context of handsfree radiotelephony.** P. Scalart, A. Benamar.
- [6] **Noise spectra** CCITT blue book, 1988, Volume 5. Supplement n°13.
- [7] **Binaural Measurements of loudness as a parameter in the evaluation of sound quality in automobiles.** Greg Michel, Gordon Ebbitt, Bruel and Kjaer Instruments Inc.
- [8] **Hands-free and handset mobile telephoning: simulated in situ assessment of telephonic signals and noise using HATS.** Bruel & Kjaer. CCITT COM XII-51. July 1990.
- [9] **Subjective evaluation of quality in car hands-free radiotelephone situation.** France Télécom/CNET. CCITT SG XII.,D.95, Brasilia, September 1991.

- [10] **Subjective evaluation of quality of communications in car hands-free radiotelephone context.** UIT-T Com 12-12, March 1993. France.

---

## B.3 References from subclause 4.2.1

- [1] **State of art in Acoustic Echo Cancellation** .A. Gilloire, E. Moulines, D.Slosk, P. Duhamel. Published in »Digital signal processing in telecommunications »

---

## B.4 References from subclause 4.3.1

"Sound Fields Around the Head/Handset System And Their Exploitation For Pressure Gradient Noise Cancellation": M.P. Hollier, Institute Of Acoustics Proceedings Vol. 12, Part 10 (1990).

"Communications in noise - a performance ranking metric", Hollier M P, Guard D R, Hawksford M O, BT Technology Journal, Vol. 10, N° 4, October 1992.

'Outline Proposal For A Noise Rejection Specification For 1/2-Rate GSM' M. Hollier, BTL, Tdoc SMG-02 03.50 32/95.

"Echo Loss Of GSM Mobiles"; I. Goetz, SMG-02 03.50 Tdoc 4/93.

"Speech & Speech Model Signals: A Comparison" ITU-TS SGXII Q7/12 & 13/12 Contribution 14, March 1993.

---

## B.5 References from subclause 4.3.3

- [1] **"The Ultimate Test Set"**, RACE R2072 WP2.2 Contribution, PTT Research, Holland.
- [2] **"Status of the Standardization activities on Acoustic Echo Control"**, A. Gilloire, Signal Processing, n°27, pp.273-279, 1992.
- [3] **"Performance Evaluation of Acoustic Echo Control: Required Values and Measurements Procedures"**, A. Gilloire, 3rd Int. Workshop on Acoustic Echo Control, Sep. 93, Lannion-France.
- [4] **"EVALUATION METHODOLOGY FOR FULL-DUPLEX ACOUSTIC ECHO CONTROLLERS"** Contribution ITU-T COM 12-40 (September 1994)), Source: FREETEL Project.
- [5] **"FREETEL FINAL REPORT "**, ESPRIT IV Project n° 6166.

---

## Annex C: Document change history

SPEC	SMG#	CR	PHASE	VERS	NEW_VERSIONS	SUBJECT
03.58	s23	new	R97	2.0.0	5.0.0	Handsfree test methods
03.58	s26		R97	5.0.0	6.0.0	R97 version change
03.58	s26		R97	6.0.0	6.0.1	Editorial changes for publication
03.58	s29		R98		7.0.0	Specification version upgrade to Release 1998 version 7.0.0
03.58	S31		R99		8.0.0	Specification version upgrade to Release 1999 version 8.0.0

---

# History

<b>Document history</b>		
V7.0.0	August 1999	Publication
V.8.0.0	March 2000	Version for Release 1999

ISBN  
Dépôt légal :