Speech enhancement in mobile devices

Anders Eriksson
Multimedia Technology Department
Ericsson Research
Stockholm - Sweden

ERICSSON
TAKING YOU FORWARD

Outline

- System overview
  - What functionality is implemented and what performance can be expected
- State-of-the-art performance of mobile devices/terminals
  - Where are the possible bottlenecks
- Challenges for the future

Overview of speech quality

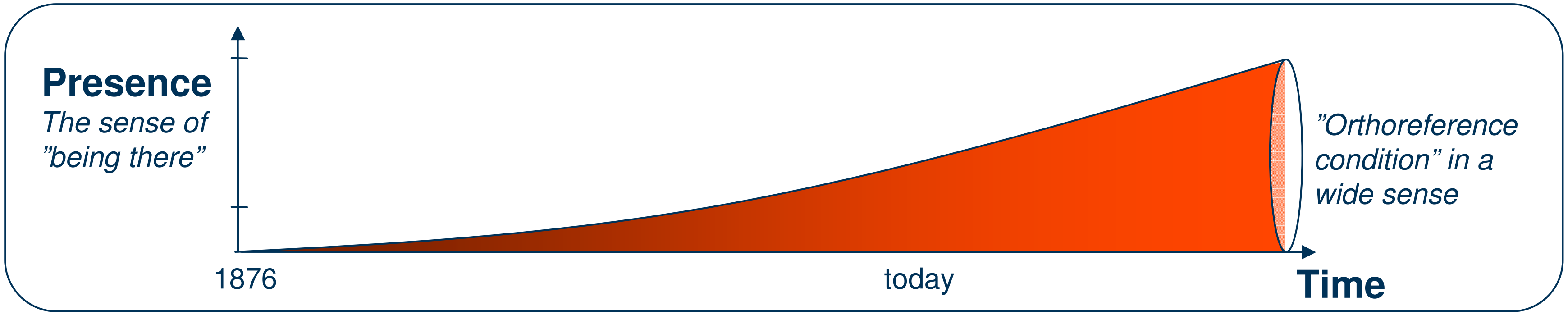
Speech quality may be defined by the ”Orthoreference condition”*:

In the ”orthoreference” communication condition, a talker and a listener communicate by speech, face to face, one meter apart in a quiet, approximately anaechoic environment. An ideal telephone system may then be defined as a system that produces the same perceived sound impression on the listener’s side as in the orthoreference condition.

* Used by ITU-T for determining gain and frequency response

The sense of ”being there”

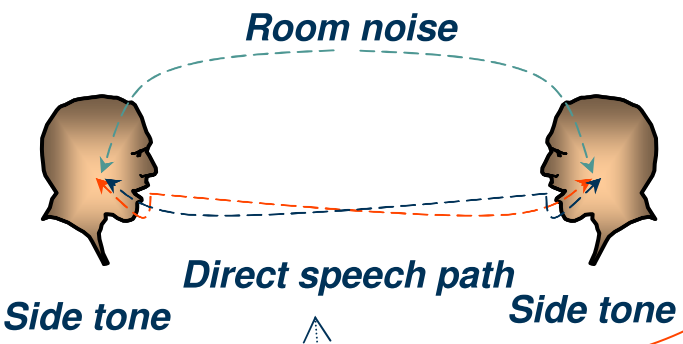
Evolutionary process of speech quality:



Speech quality in a mobile network

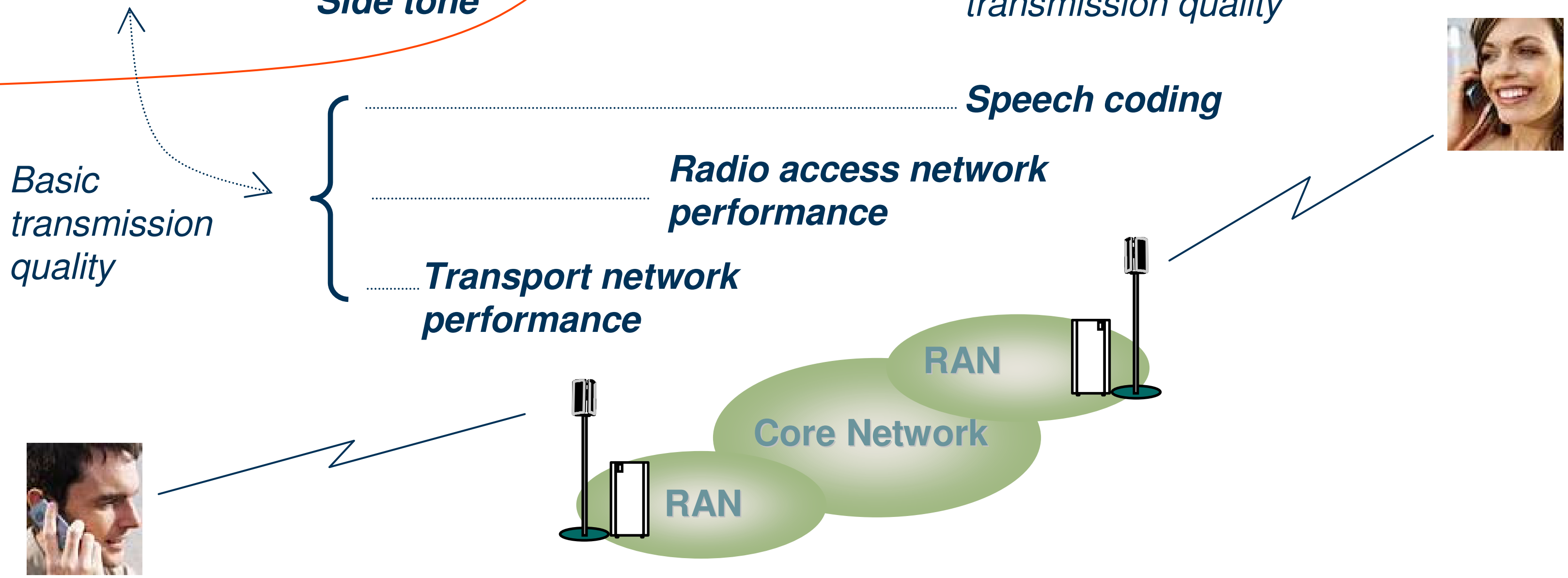
Areas affecting the speech quality

Face to face conversation



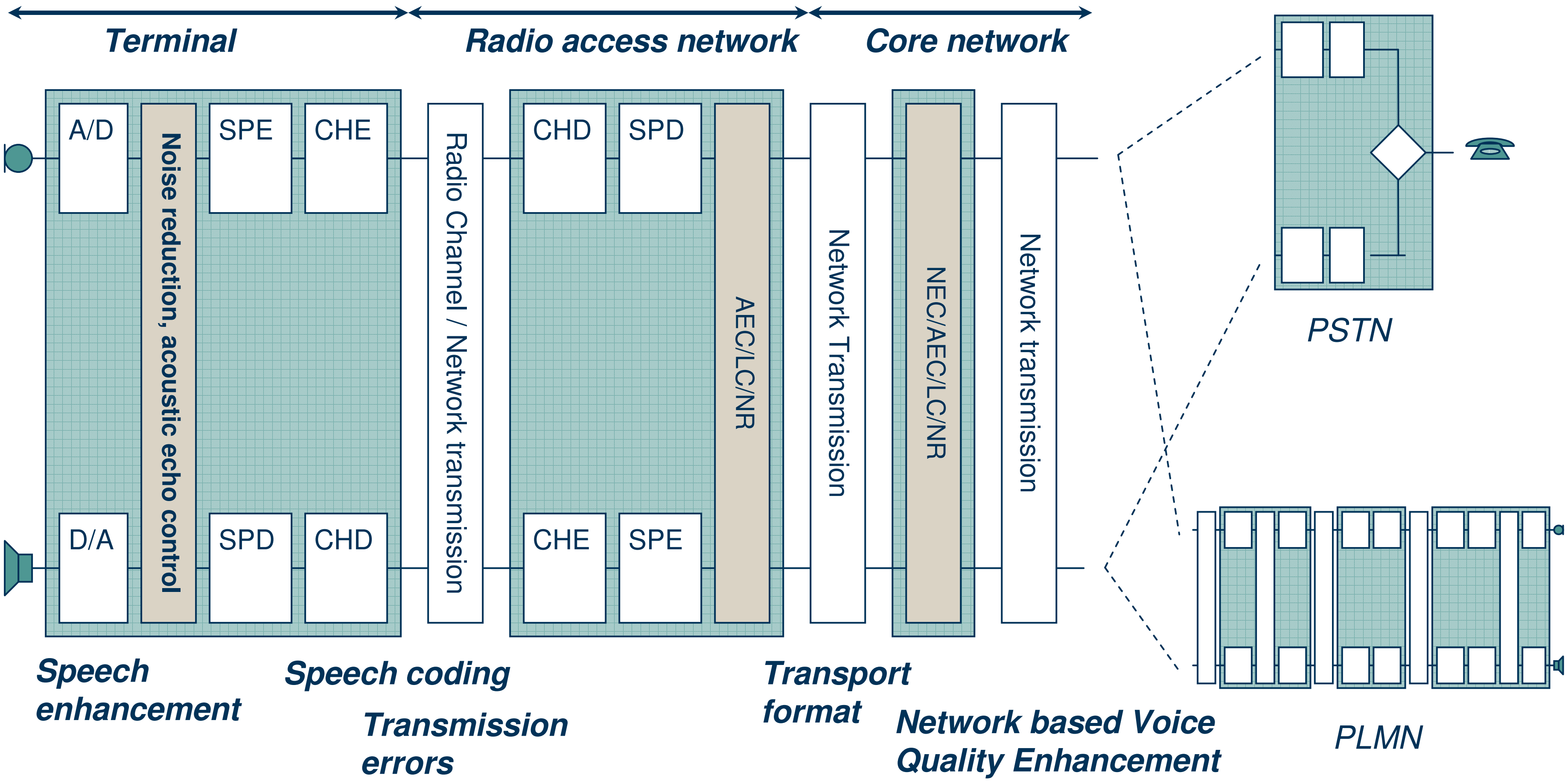
Conversation over mobile telephone connection

Terminal:
Optimize performance with respect to capturing and rendering of speech in order to replicate a face to face conversation and utilize the transmission quality



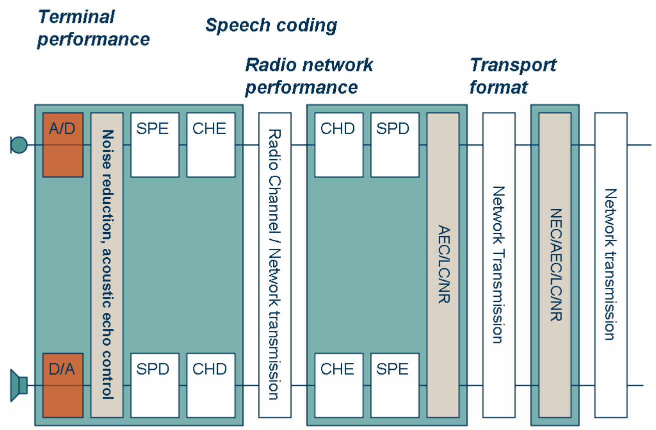
System overview

Processing elements in the transmission chain that affect the speech quality



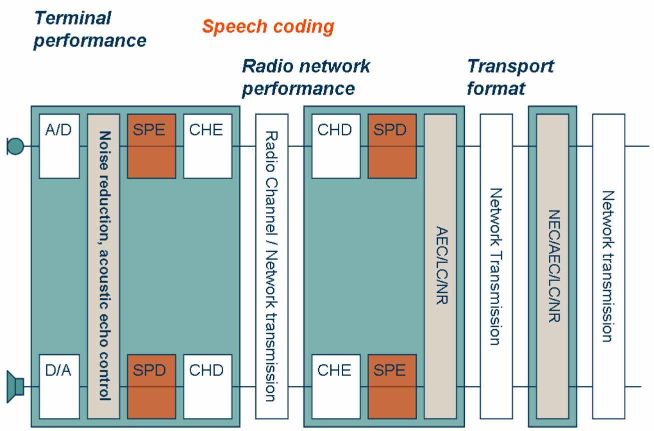
Digitization

Speech and audio bandwidth

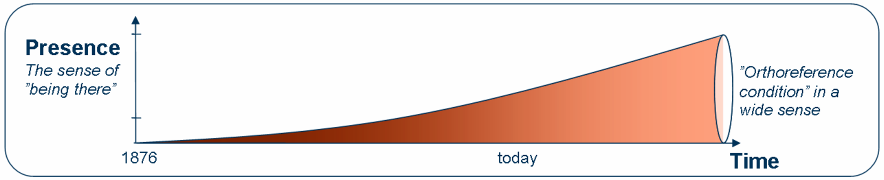


System	Bandwidth [Hz]	Sample rate [Hz]	Resolution [bits]	Bit rate [kbit/s]	Coded bit rate [kbit/s]
Telephony, narrowband	200-3800	8000	12-13	94-104	4-64
Wideband telephony	50-7000	16000	14-16	224-256	6-64
Audio	20-20000	44100, 48000	16-24	705-1152	24-196

Speech coding

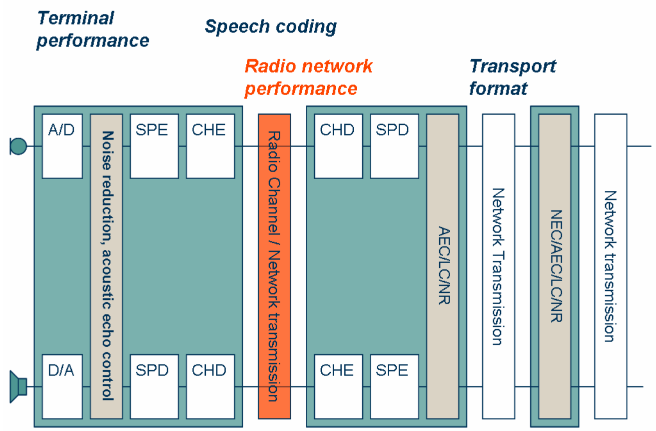


- The speech encoder maps blocks of samples of uniform 14-16 bit PCM format to encoded blocks
  - Enhanced system capacity due to reduced source bit rate
  - Improved robustness against transmission errors
  - Block size usually corresponding to 20 ms of speech for low transmission delay
- The speech decoder maps encoded blocks to uniform 16 bit PCM format
- The Adaptive Multirate (AMR) codec was standardized in 1998 by ETSI for the GSM system and was later adopted for the WCDMA 3G system
  - Operating on 8 kHz sampled speech with 8 source rates between 4.75 kbit/s and 12.2 kbit/s + low rate background noise encoding mode (DTX)
  - Highest source rate gives a speech quality similar to 64 kbit/s G.711 PCM (used in the fixed telephony network)
- The Wideband AMR codec was standardized in 2000 by ETSI/3GPP
  - Operating on 16 kHz sampled speech with 9 source rates between 6.60 kbit/s and 23.85 kbit/s + low rate background noise encoding mode (DTX)
  - Wideband (16 kHz sampled) speech gives increased speech quality

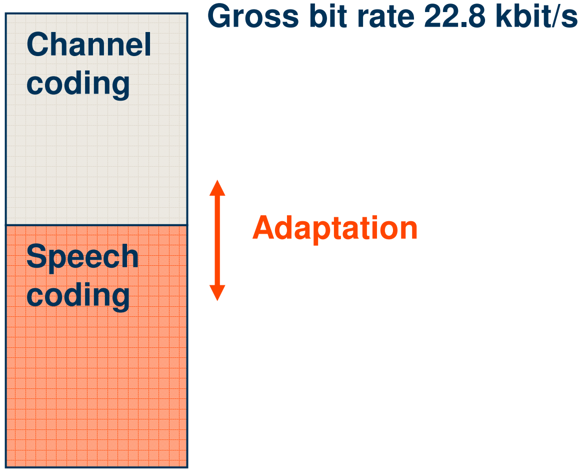


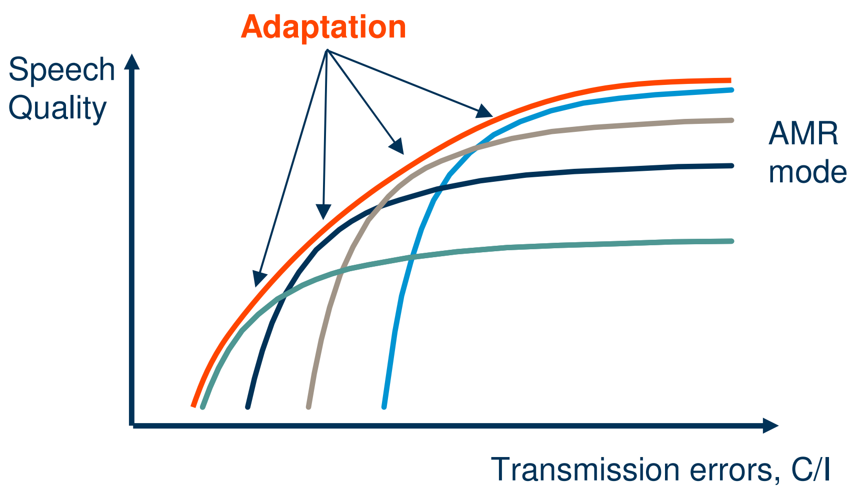
Radio network

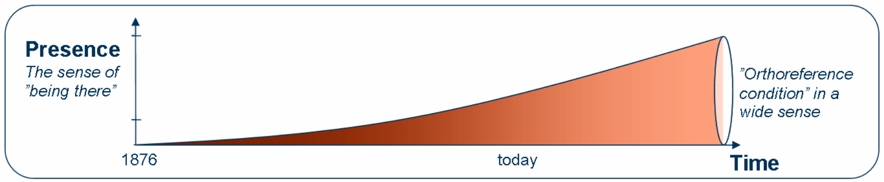
Robustness to transmission errors



- Radio transmission errors introduces loss of speech frames
- AMR gives increased robustness by trading source bit rate against channel coding



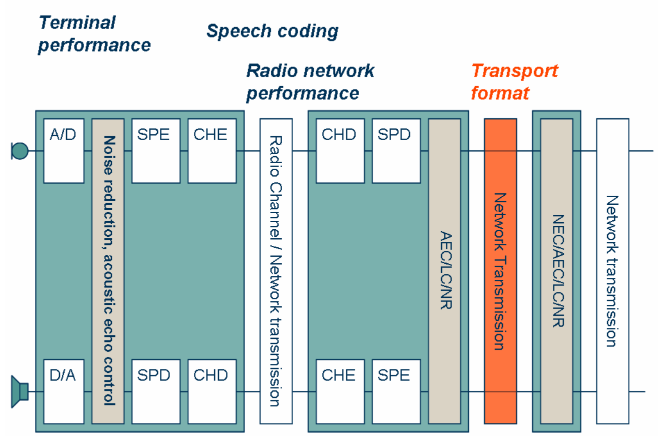


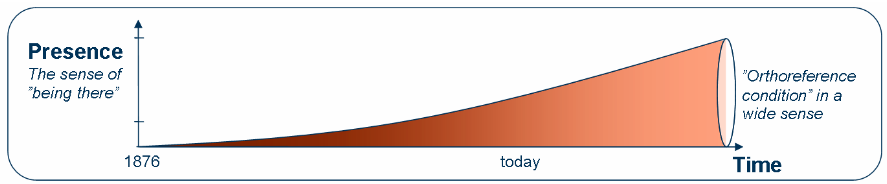


Core network

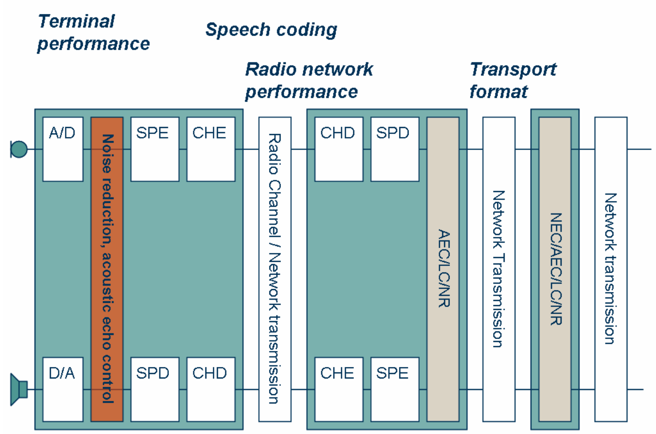
Tandem and transcoder free operation

- ”No” transmission errors in the fixed circuit switched transport network
- Slight loss in speech quality due to multiple speech encodings if transmission using 64 kbit/s PCM in the core network
  - Improve speech quality by transmitting AMR coded speech also in the core network

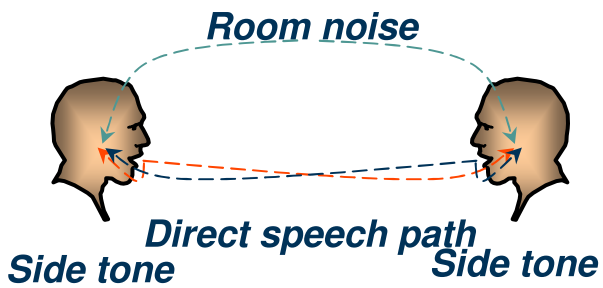


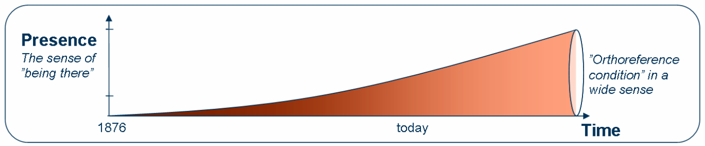


Terminal design



- The mobile transmission network (“the direct speech path”) allows for a speech quality similar to 16 bit uniform PCM
  - Wideband speech coding will further increase the intrinsic speech quality
- The speech signal that is rendered in the terminal should have properties similar to a face-to-face conversation
  - Proper side tone to adjust your own voice level
  - Pleasant speech and noise level of the receiving speech
  - No echo of your own speech
- According to the GSM/UMTS standard it is the responsibility of the sending terminal to perform proper signal conditioning (“speech enhancement”)



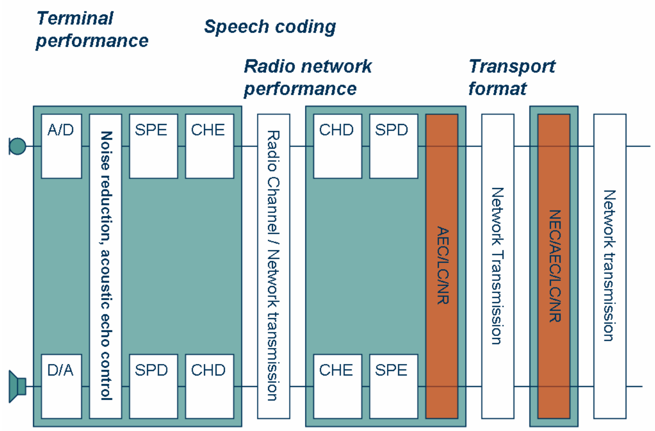


Requirements on terminals

- Requirements on the acoustic properties of terminals in Technical Specification 3G TS 26.131 ”Terminal Acoustics Characteristics for Telephony”
  - Send and receive speech signal levels
    - The electro-acoustic losses of the terminal should be within specific limits to give good speech quality and allow interoperability between different terminal vendors
  - Noise level
    - Strong background noise has an adverse effect on speech coding
    - Listener comfort
  - Acoustic echo
    - Due to the transmission delay (~100 ms/link) the presence of echo would severely impact the communication
- The requirement holds for all operating modes of the terminal
  - Handset, speaker mode, car handsfree, etc.
- Speech enhancement algorithms are most often needed to fulfill the requirements
  - General rule to apply speech enhancement as close as possible to the source
- Requirements on low algorithmic delay in 3G TS 43.005 ”Technical performance objectives”
  - No room for extra algorthmic delay in the total delay budget

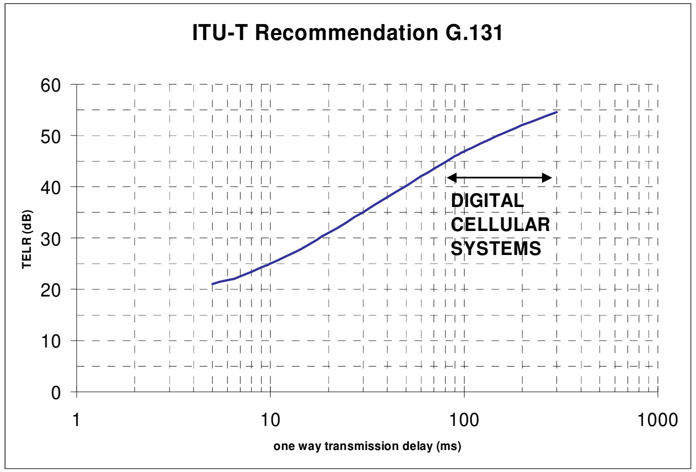
Network Based VQE

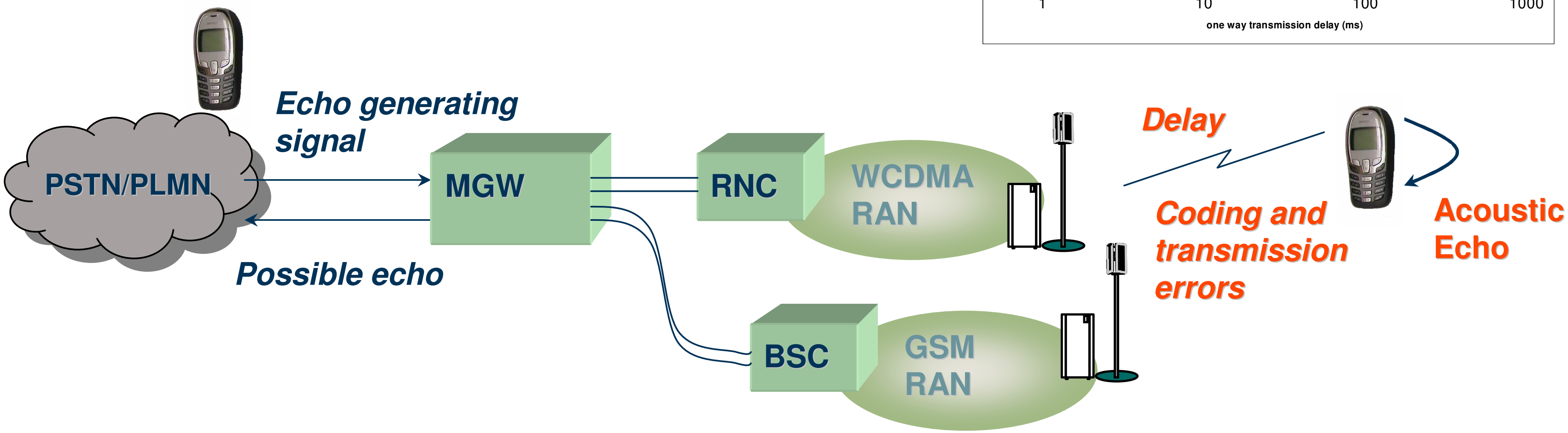
Voice Quality Enhancement



- Speech enhancement elements introduced in the transport network
- Gives improvement on the perceived overall quality of the speech transmission *(“the speech quality”)* in three aspects
  - Speech level *(Level Control)*
  - Noise level *(Noise Reduction)*
  - Acoustic echo *(Acoustic Echo Control)*
- The need is linked to the terminal behaviour
  - Terminals designed according to the specifications would not require VQE
  - A few terminals (mainly based on old platform design) may not meet the spirit of the specifications and may benefit from VQE
- Used for *correction* and to a limited extent enhancement of the speech quality
  - Can only correct the speech quality to a level that is significantly below what is achieved with good terminal design
  - Should give benefits in situations for which the terminal performance is not adequate
  - Should not degrade the quality in situations where the terminal performance is good
- The limit on the speech quality is set by the terminal performance, radio network performance and speech coding
  - VQE can not correct poor radio network performance or speech coding performance

Acoustic echo
System aspects

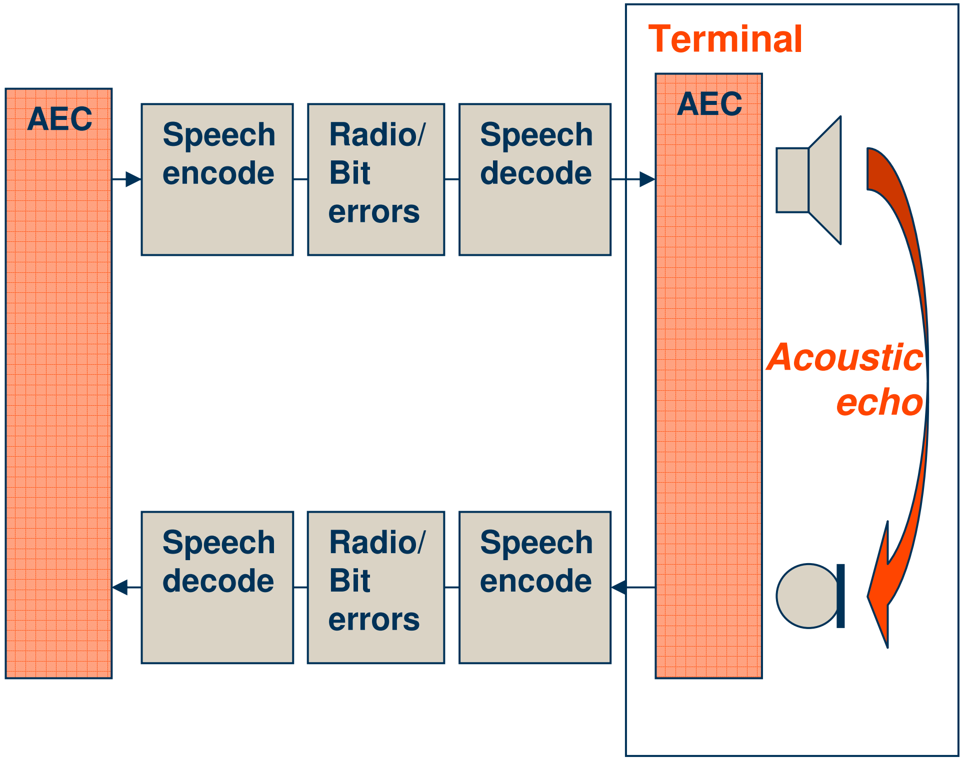




- Acoustic echo in combination with delay can be annoying to the remote talker (see e.g. ITU-T Recommendation G.131)
- Acoustic echo will increase the voice activity factor of DTX in the up-link and thus reduce the radio efficiency
- Echo characteristics
  - Highly non-linear echo with long delay due to speech coding/radio transmission

Control of acoustic echo
System aspects

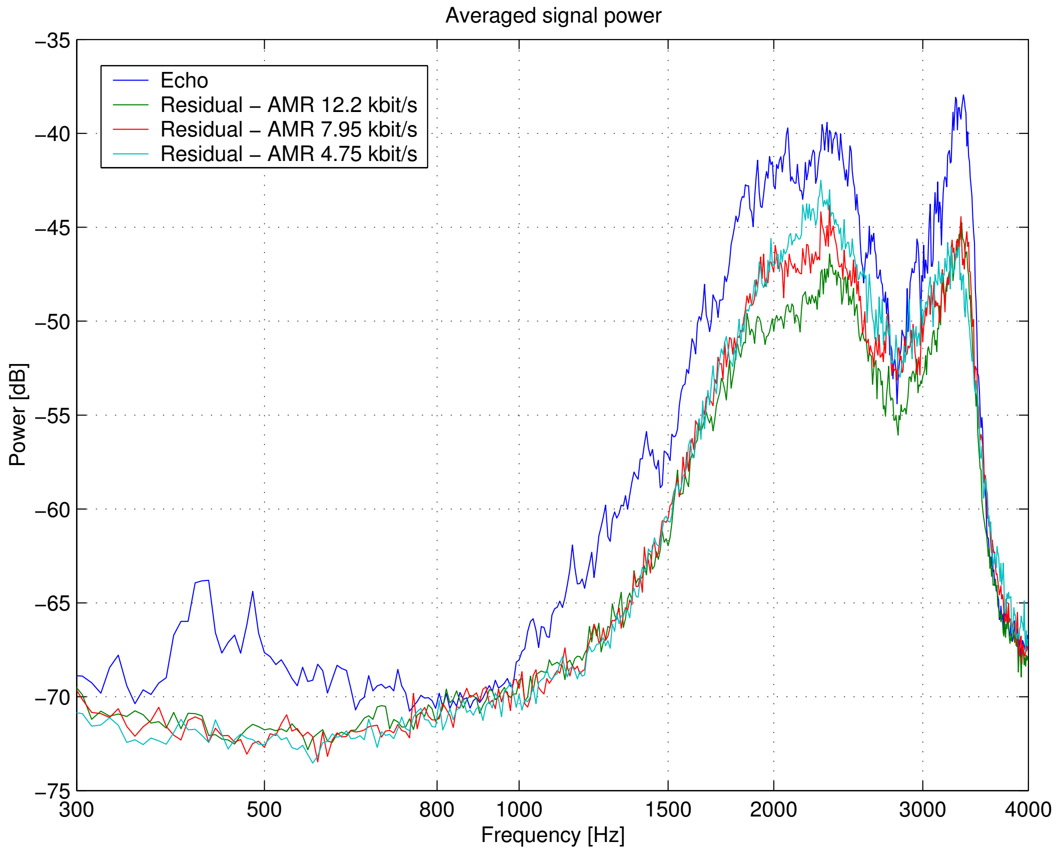
- Acoustic Echo Canceller in the mobile terminal
  - Fairly linear echo path
  - No echo path delay
  - A priori knowledge of the echo characteristics
- Acoustic Echo Canceller in the transmission network
  - Speech coding/transmission errors reduces the linearity
  - Rather unknown echo path delay due to transmission
  - No knowledge of terminal characteristics
- Performance and network capacity calls for AEC in the terminal



Speech coding in the echo path

Limit the linear echo attenuation by the AEC

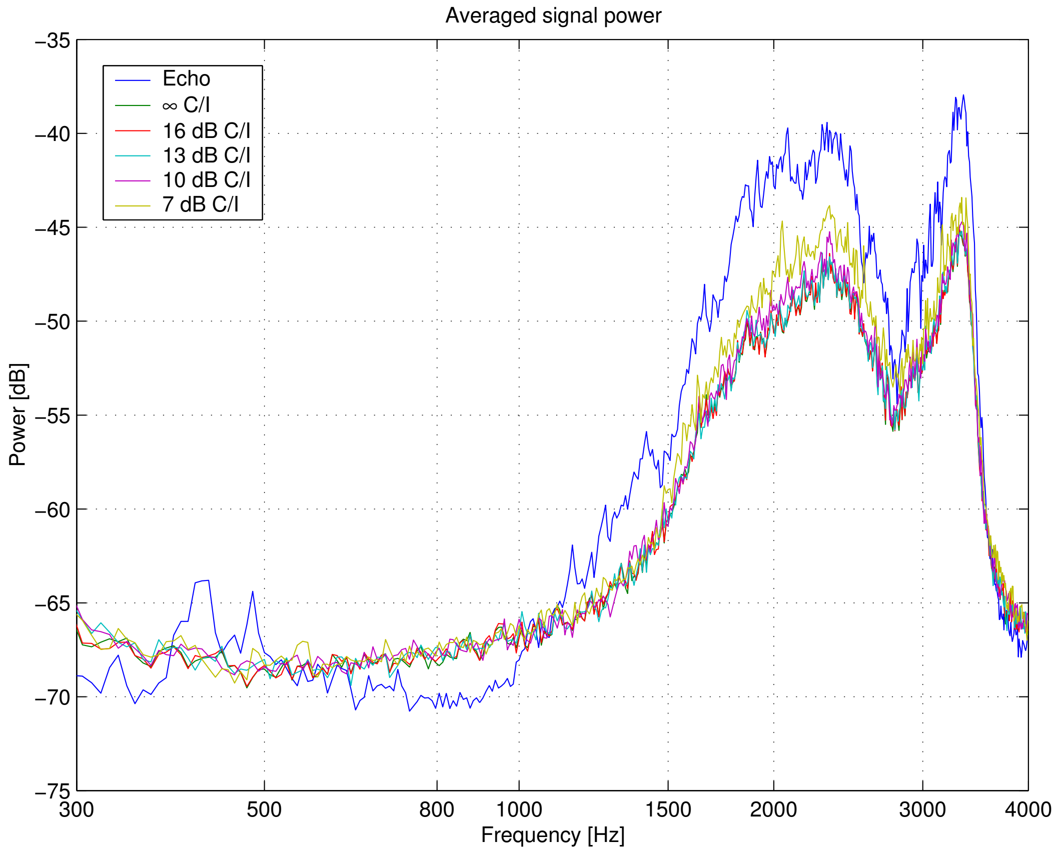
- **AMR 12.2 / EFR**
  => 0 - 10 dB echo reduction
- **AMR lower bit rates**
  => 0 – 5 dB echo reduction



Transmission errors in the echo path

Limit the linear echo reduction by the AEC

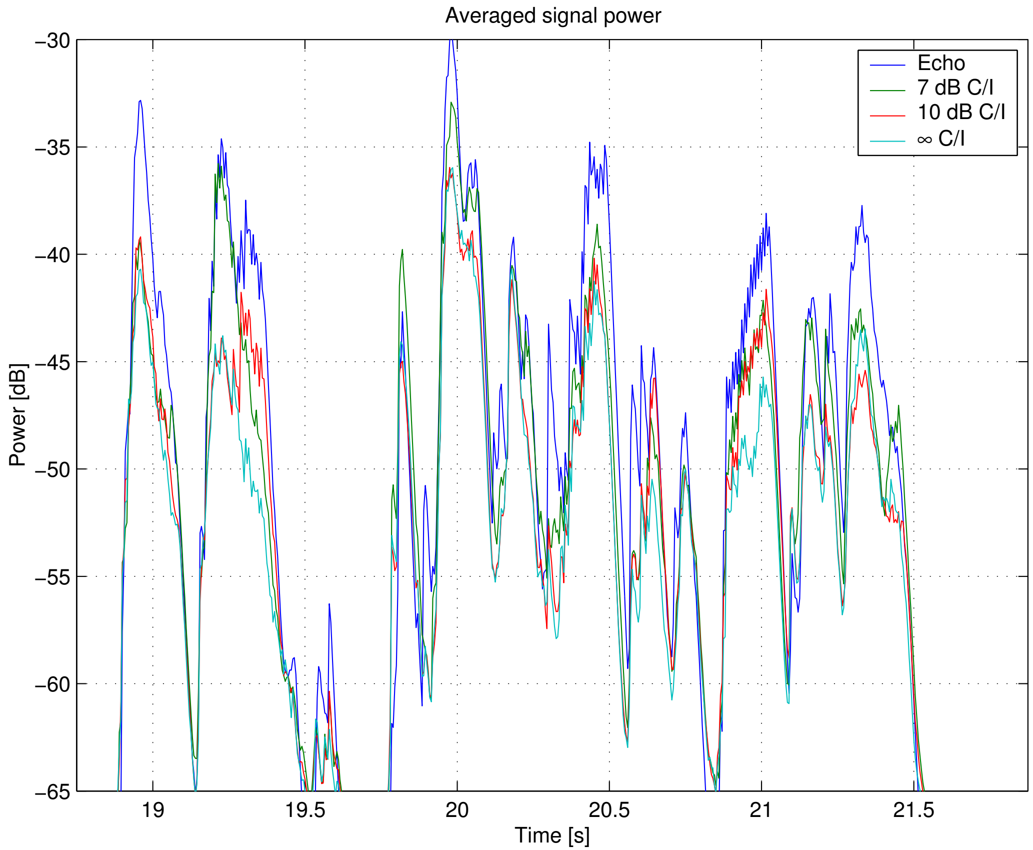
- **Negative effects at 10 dB up-link C/I (0.3 % frame error rate)**
- **Severe problems below 7 dB up-link C/I (4 % frame error rate)**



Transmission errors in the echo path

Limit the linear echo reduction by the AEC

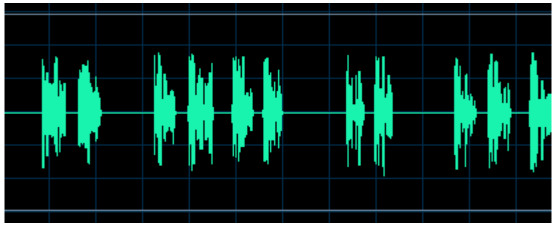
- **Time varying disturbance**



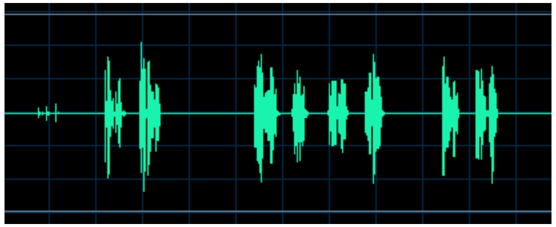
Echo impact on DTX/VAF

Speech signal diagrams

Downlink speech

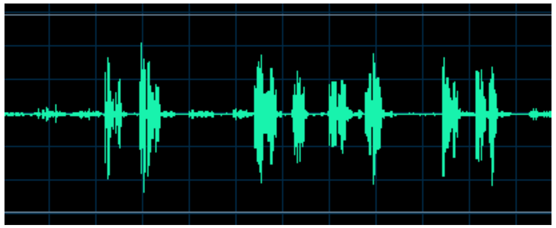


Near end speech



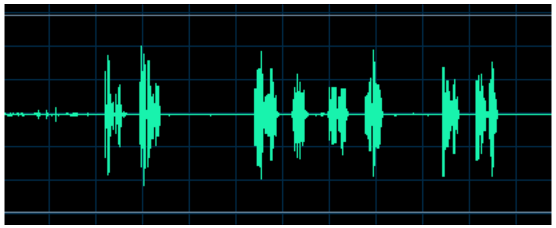
This is the ideal case: (no echo at all), resulting in lowest possible uplink activity without any loss in speech

Terminal input (microphone input speech with echo)



Although the echo is low it triggers the VAD quite often andgenerates a high uplink radio activity

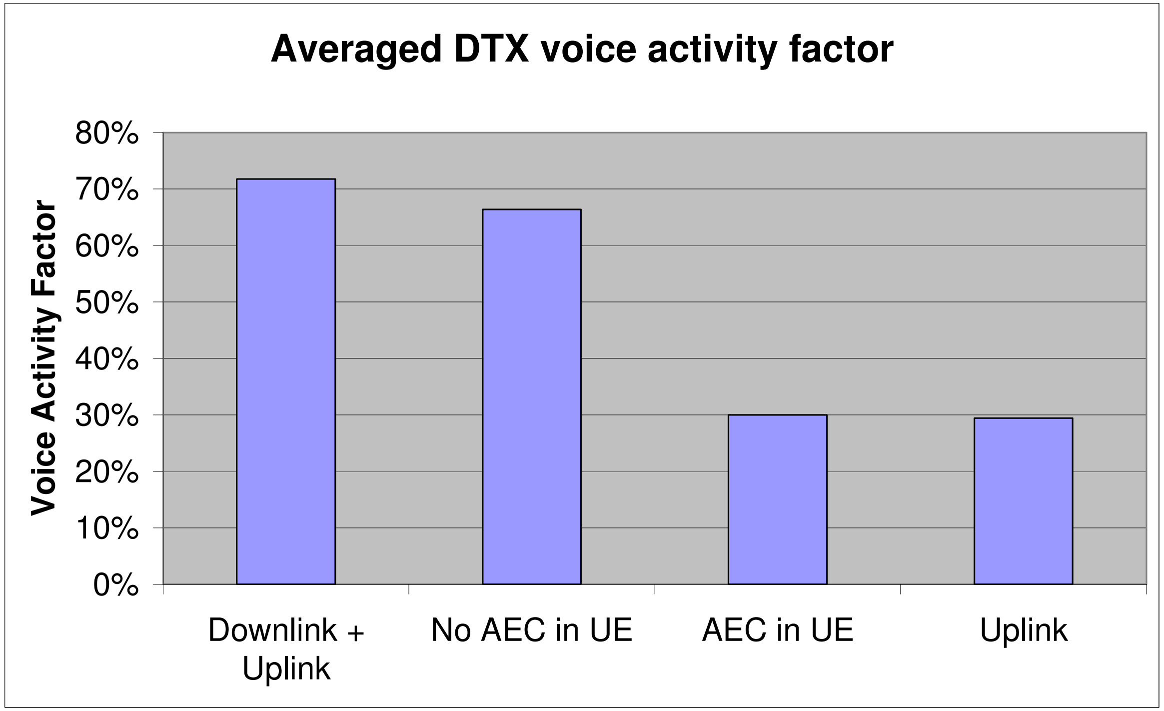
AEC output (uplink speech with terminal AEC)



The VAD activity here is close to the ideal case (no echo at all)
An AEC in the network can not reduce the unnecessary voice activity due to echo

Echo impact on DTX/VAF

DTX: Discontinuous Transmission in GERAN = **SCR:** Source Controlled Rate in UTRAN
VAF: Voice Activity Factor = directly related to the Radio Activity Factor



Downlink+Uplink: the activity of the sum of both speech signals (for reference only)

No AEC in UE: the activity in the uplink with terminal echo and no AEC in the UE

AEC in UE: the activity in the uplink with terminal AEC, i.e. the near end user's speech

Uplink: the air interface activity for an echo free terminal (ideally near end speech only)

Stickphone, i.e. not a clam-shell phone

2 volume settings (max, max – 3)

2 gender (female, male)

3 downlink speech levels (-26 dBov, -20 dBov, -16 dBov)

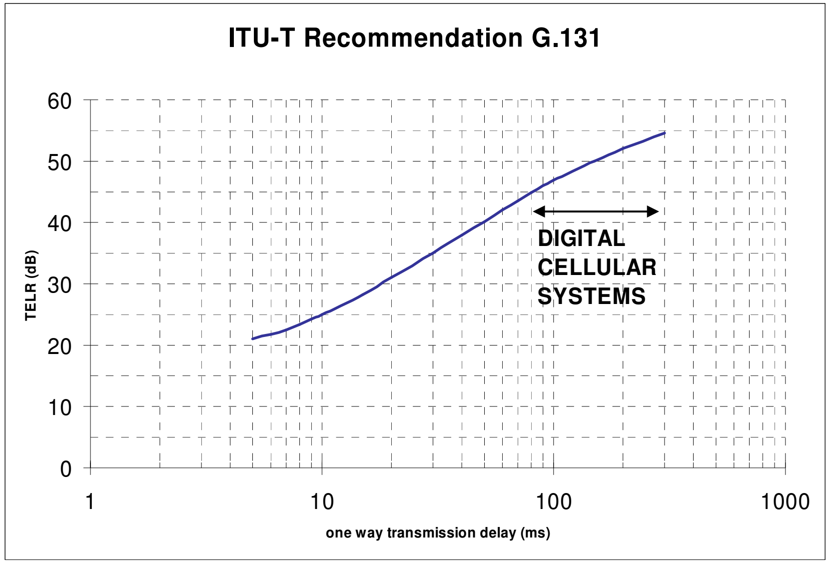
4 uplink speech levels (silence, 84 dB SPL, 90 dB SPL, 96 dB SPL)

Acoustic echo
System requirements

- Requirement specification 3GPP TS 26.131 on terminals:

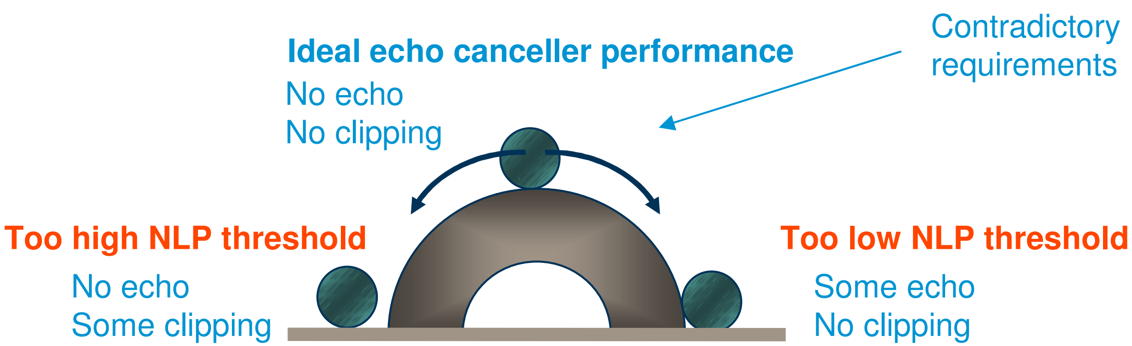
  "The echo loss presented by the 3G/GSM network at the POI should be at least 46 dB during single talk. This value takes into account the fact that UE/MS is likely to be used in a wide range of noise environments." c.f. ITU-T Recommendation G.131

- Achievable via
  - Acoustic design of the terminal
  - Signal processing (Acoustic Echo Cancellation) of the microphone signal before speech coding/up-link transmission
    - Most/all terminals of today include acoustic echo cancellers to allow for a flexible industrial design
- Echo may still be noticeable due to
  - Inferior design of echo canceller
  - Large variations of the echo path characteristics



Echo canceller performance
Subjective requirements

- Echo reduction is obtained by a combination of linear echo reduction and residual echo suppression techniques
- To a certain extent two contradictory requirements that needs to be balanced when judging the performance
  - No **echo** (residual echo not handled by the echo canceller)
  - No **clipping** (loss or distortion of speech or background sound from the near end)
- Subjective evaluations
  - Stress the duplex nature of the echo canceller: both far-end and near end signals are needed to evaluate performance
  - Masking effects from the side-tone difficult to take into account: use both listening tests and conversational tests



Survey of terminal echo situation

Performed in early 2006 on state-of-the-art terminals

- Evaluated the echo performance in handheld and speaker phone mode in both live conversation and a controlled environment using artificial head and recorded speech material
- All terminals in this survey (19 terminals) have acceptable or good speech quality performance with respect to echo

Noise Reduction

Overview

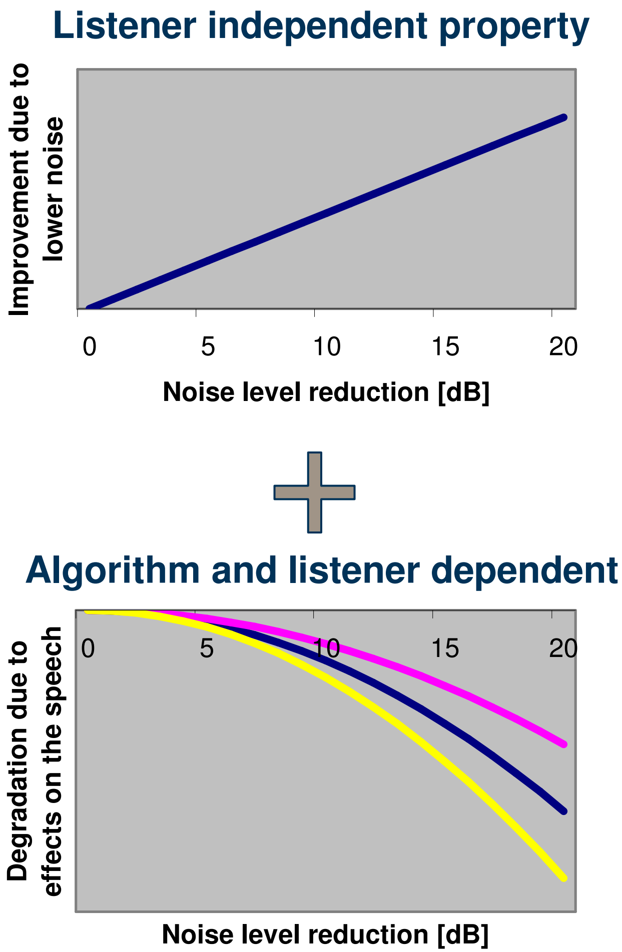
- A high background acoustic noise level is annoying to the listener side
  - Listener fatigue (“the ears get tired”)
  - Difficulties to understand each other
- Background noise is a natural part of a conversation
  - Provides information about the surrounding environment of the person we talk to
- Reduce the noise level in the speech signal to a comfortable level but retain the basic characteristic of the noise
  - Due to the design of terminals the SNR is positive in most situations
  - Not primarily for speech intelligibility
- Does *not* reduce the background noise for the mobile user in a noisy environment

Requirements on Noise Reduction

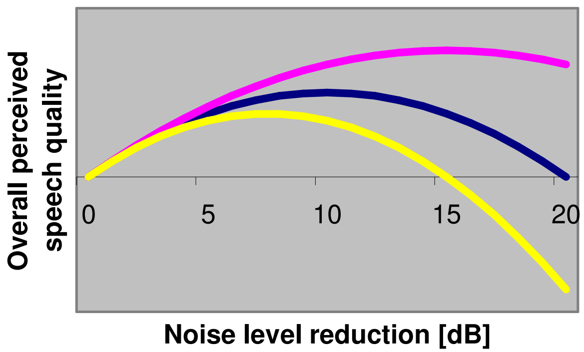
Terminal design

- UMTS/GSM requirements on the terminals in mobile networks on their relative sensitivity to the talkers voice and ambient background noise
  - 3GPP TS 26.131 – "Terminal Acoustic Characteristics for Telephony"
    - At least 0 dB single figure DELSM, +3 dB recommended
- Achievable via
  - Acoustic design of the terminal
  - Signal processing (Noise Reduction) of the microphone signal before speech coding/up-link transmission

Evaluation of the perceived overall quality of Noise Reduced signals



- Two dimensional problem
  - Improvement due to lower noise level
    - Fairly listener independent
  - Degradation due to effect on speech
    - Listener dependent
    - Algorithm and coder dependent
- Perceived overall quality (noise reduction vs. impact on speech) is very dependent on listeners



Summary of speech quality from mobile terminals

- Advances in DSP technology and increased focus on speech quality has lead to improved terminal performance with respect to echo and speech and noise level
- New use cases and wideband speech will give a continued interest for enhanced speech quality

Challenges

Processor capacity and algorithm complexity

- Moore's law
  - Processing power will continue to increase at even pace
- Battery life time needs to match the demands on processing consumption
  - Talk-time is a top of the list feature
  - Fixed point implementation will still be important for reduced power consumption
- Strong requirements on limited computational complexity of speech enhancement algorithms will prevail for a foreseeable future
  - Introduction of wideband will demand higher complexity for the same functionality
  - Possibility for better utilization of processor capacity by exploiting parallelism and vector processing in algorithm design

Challanges

Advances in transport and supplementary technology

- Wideband, 16 kHz sampled speech
  - New demands on both performance and complexity
  - ”The same” algorithms takes 2-4 times more complexity
  - Statistics of speech signal even more intriguing
- Synthetic stereo
  - Java specification JSR 234 Advanced Multimedia Supplements: 3D audio used for enhanced presentation
- IP Multimedia Subsystem (IMS), Mobile Telephony Service over IMS (MTSI/MMTel)
  - Framework for VoIP with operability and quality-of-service
- Generic Access Network
  - Mobile at home via Bluetooth or Wi-Fi/802.11

Challanges
End-user aspects

- Consumer electronic - Industrial design is #1 priority
  - Versatility – The same algorithm for many designs
  - Robustness – Benefits and no degradation
- Use cases
  - Handset
  - Portable handsfree
  - Speaker mode
    - Video telephony
    - Group communication
    - Car handsfree
  - Advanced conferencing and microphone arrangements?
- Loudspeaker enhancements
  - Benefits for the own user

Acronyms and abbreviations

AEC	Acoustic echo canceller
A/D	Analogue to digital conversion
BSC	Base station controller
CHD	Radio channel decoding
CHE	Radio channel encoding
D/A	Digital to analogue conversion
dB A	Acoustic sound pressure, A-weighted
dBov	Electric signal level relative digital overload
FLC	Fixed level compemsation
LC	Level compensation
MGW	Media gateway
NLC	Noise level compensation
NR	Noise reduction
PLMN	Public land mobile network
PSTN	Public switched telephone network
RAN	Radio acces network
RNC	Radio network controller
SPD	Speech decoding
SPE	Speech decoding
VQE	Voice quality enhancement

ERICSSON

TAKING YOU FORWARD