Wireless Speech and Audio Communications
A Time Warp

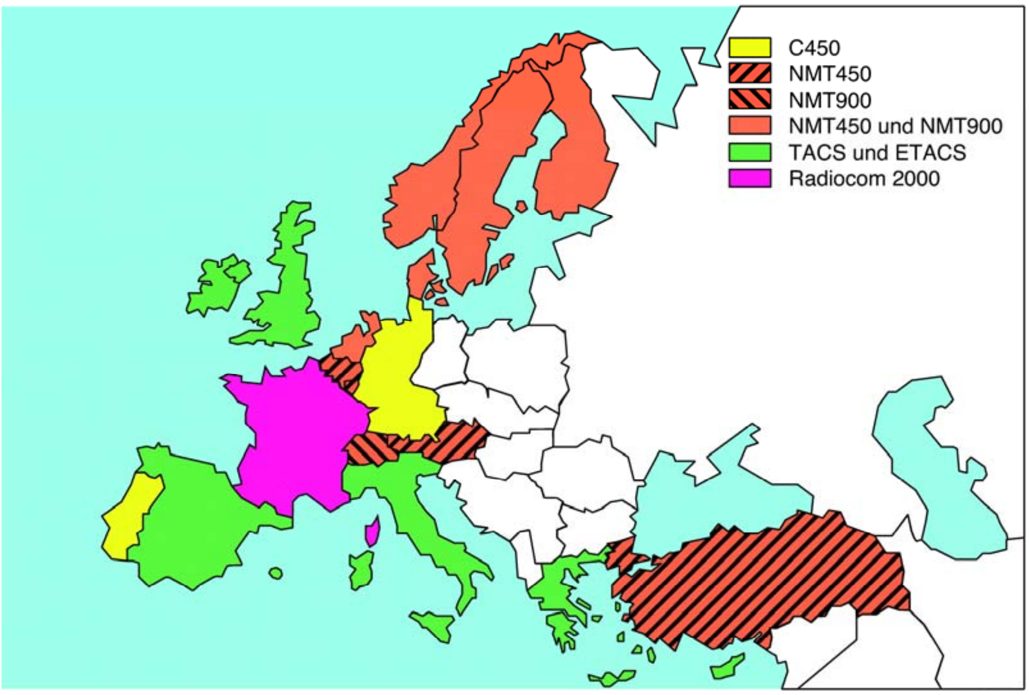
Peter Vary

EUSIPCO, 1.9.2015, Nice

Audio examples will be made available at: http://www.ind.rwth-aachen.de/en/publications/

Time Warp Prologue | 1985

- Non compatible analog cellular standards in Europe



Milestones

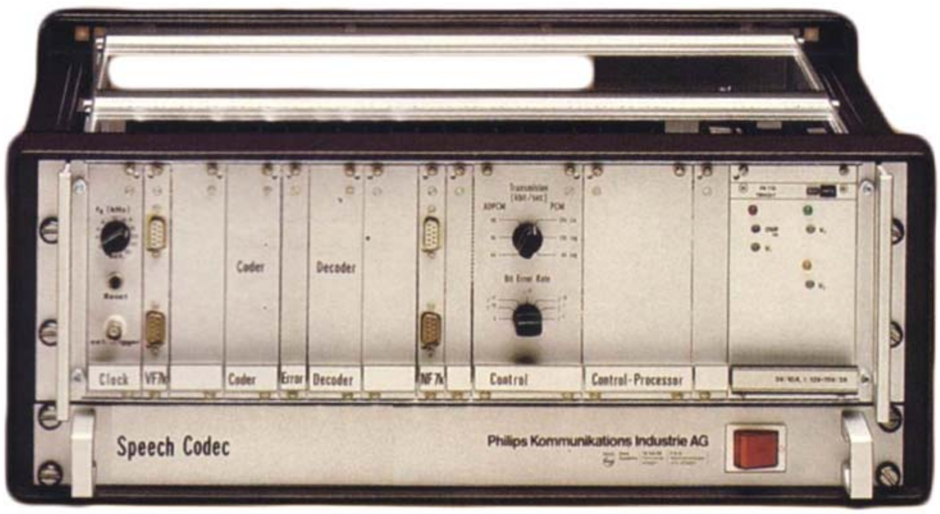
1984 | French-German Initiative for Digital European Cellular Radio

1988 | GSM Standard: Global System for Mobile Communications

1990 | European IP-Backbone-Network EBONE

1992 | Commercial GSM Networks

Speech Codec | 1985



Karl Hellwig | 1985

GSM Mobile Station | 1989

RWTHAACHEN

First Hand-Held GSM Mobile Phone | 1992

Motorola International 3200, „The Brick Phone"

- ca. 2.500 €
- 750 mAh battery
- **520 grams**
- Talk time 60 minutes
- Standby **8 h**
- No data service, no SMS messaging

RWTHAACHEN

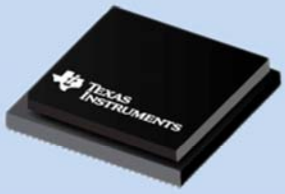
iPhone 6 | 2015

- 699 – 999 €
- **129 grams**
- Talk time **14 h** (3G)
- Standby up to **250 h**
- GSM, UMTS, LTE, 5G, WiFi, Bluetooth, GPS, NFC
- A8 processor, 64 bit architecture
- M8 motion co-processor, 2 billion transistors
- Gyro sensor, barometer, …
- Apps, apps, apps, ….

→ **The 2015 smartphone is a 1985 hand-held supercomputer!!**

30 Years of Moore´s Law | 1985 - 2015

- Evolution of DSP technology
- Doubling 15 times: $2^{15} = 32.768$

	1985 | NEC µPD 7720	2015 | TMS 320C6678 8xMulticore	Factor:
			
Clock	8.33 MHz	1.4 GHz	168
Data RAM	256 Bytes	8.45 MBytes	33000
Multiplications (fixed point)	4×10^6/s	358.4×10^9/s	89600

The Voice Quality Issue | 1992 - 2015

1992 | Mobility is the luxury, not voice quality

2015 | Voice quality will be a major issue
→ users rely more and more exclusively on mobile phones

Detrimental quality factors & countermeasures

- Quantization Noise
- Bit Errors
- Packet Losses
- Latency
- Audio Bandwidth

⇩

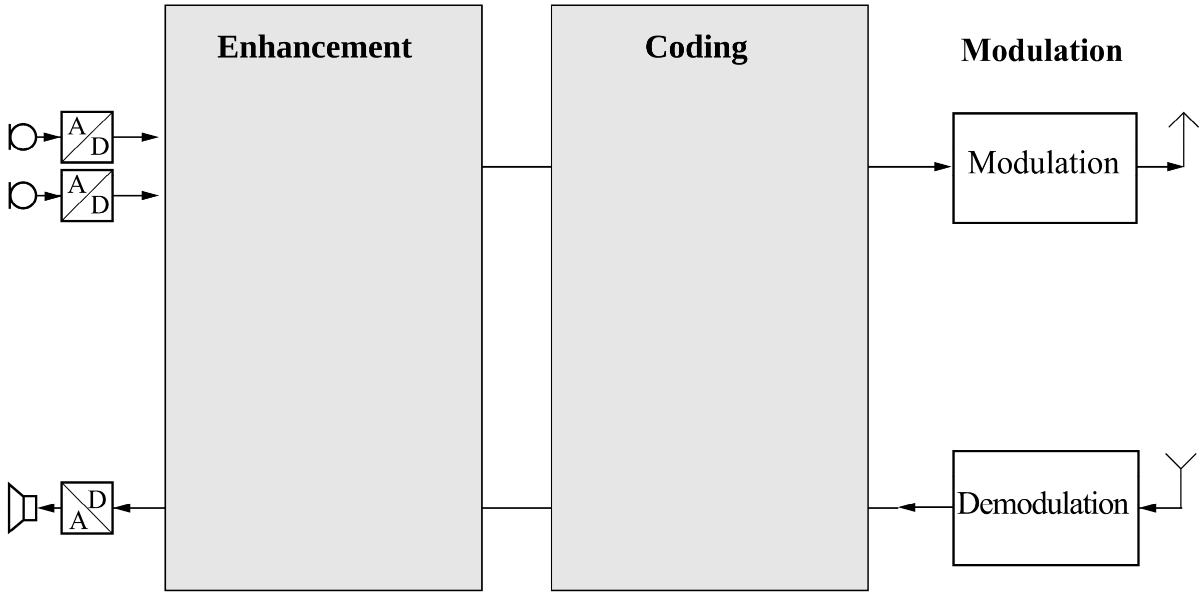
Coding

- Audio Bandwidth
- Background Noise
- Loudspeaker Echo
- Wind Noise
- Room Reverberation

⇩

Enhancement

Voice Quality Improvement | 1992 - 2015



Time Warp | 1985 – 2015

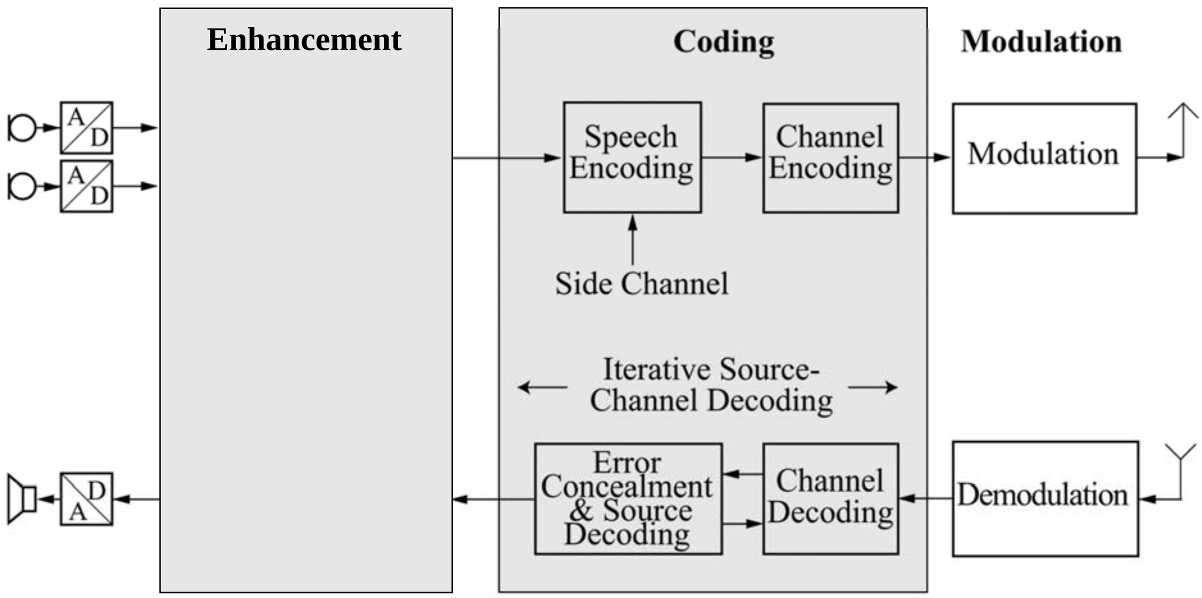
Coding

- Telephone-Voice & HD-Voice
- Steganographic Side Channel
- Error Concealment
- Joint Source-Channel Decoding

Enhancement

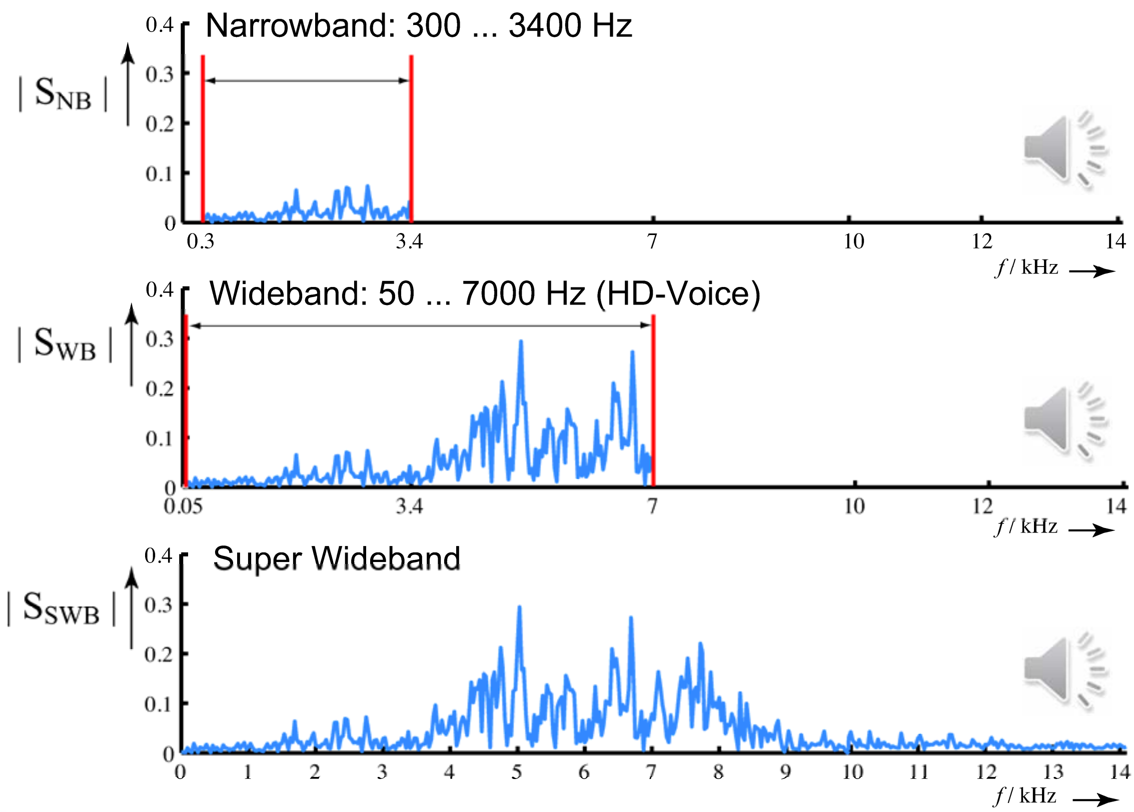
Trends

Coding in a Mobile Phone



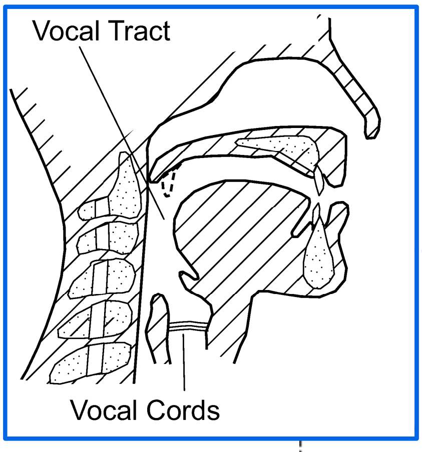
RWTHAACHEN

➢ Telephone-Voice, HD-Voice, and Beyond

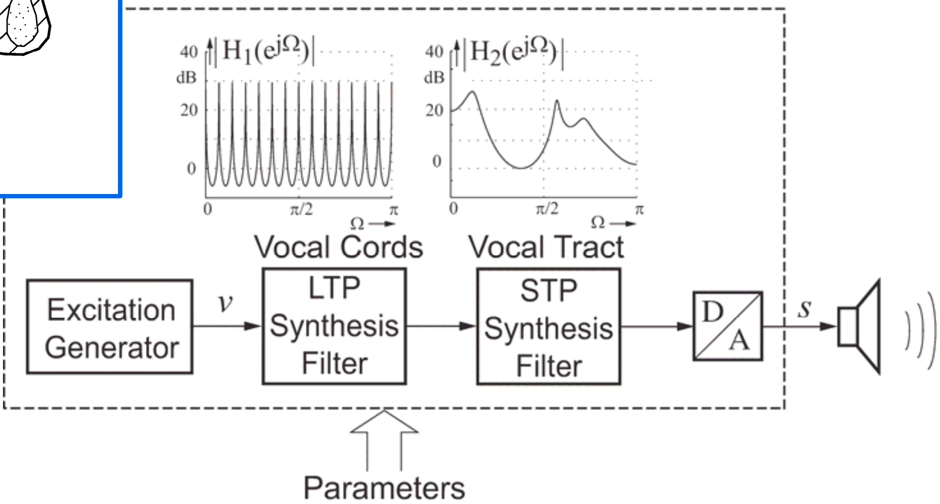


RWTHAACHEN

Model Based Speech Coding

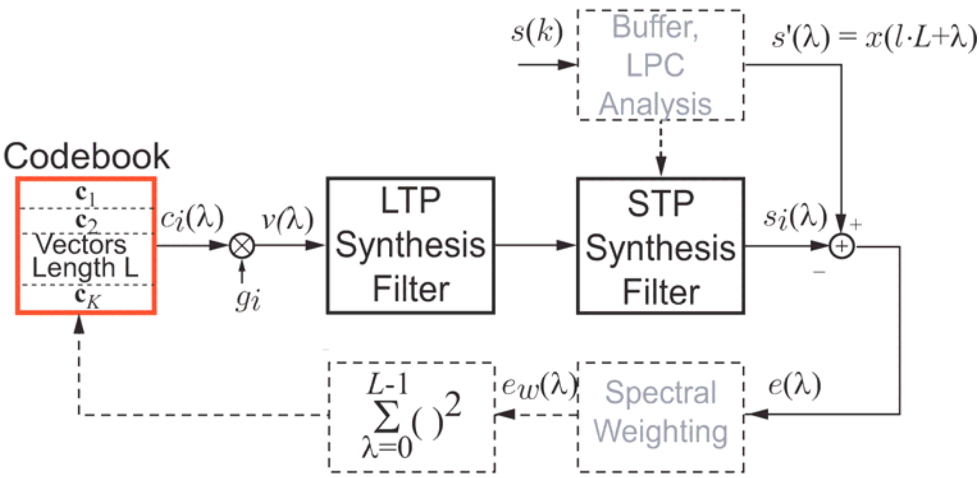


- A naturally sounding vocoder
  - 1.5 bits or less per sample (on average)
  - STP: Short Term Prediction (spectral envelope)
  - LTP: Long Term Prediction (pitch)



CELP: Code Excited Linear Prediction

- Analysis-by-synthesis coding



STP = Short Term Prediction (spectral envelope)
LTP = Long Term Prediction (pitch)

CELP: B.S. Atal, J.R. Remde | 1982
M.R. Schroeder, B.S. Atal | 1995

Speech Codecs for GSM, UMTS, LTE, and IP

	f_s/kHz	WMOPS	kbit/s
Full Rate / Half Rate Speech Codecs			
1988 | FR	8	**3.4**	13.0
1994 | HR	8	**18.5**	5.6
Adaptive Multi-Rate Speech Codecs			
1998 | AMR-NB	8	**≤ 17**	4.75 … 12.2
2001 | AMR-WB (HD)	16	**≤ 39**	6.6 … 23.85
2005 | AMR-WB+ (HD+)	32	**≤ 72**	6.6 … 32.0
IP Speech Codecs			
2006 | ITU G.729.1	8 or 16	**19 … 36**	8.0 … 32.0
2009 | ITU G.719	48	**18**	32 … 128
2012 | IETF (Opus, mono/stereo)	8 - 48	**≤ 40**	8 … 128
2015 | 3GPP EVS	8 - 48	**≤ 86**	5.9 … 128

CELP: B.S. Atal, J.R. Remde | 1982
RPE-LTP: P. Vary, J. Sluyter, C. Galand | 1988

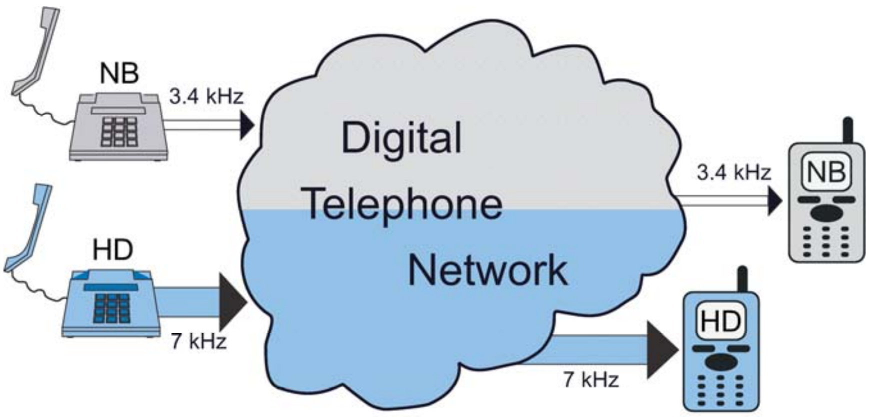
Speech Codecs for GSM, UMTS, LTE, and IP

	f_s/kHz	WMOPS	kbit/s
Full Rate / Half Rate Speech Codecs			
1988 | FR	8	**3.4**	13.0
1994 | HR	8	**18.5**	5.6
Adaptive Multi-Rate Speech Codecs			
1998 | AMR-NB	8	**≤ 17**	12.2
2001 | AMR-WB (HD)	16	**≤ 39**	23.05
2005 | AMR-WB+ (HD+)	32	**≤ 72**	24.0
IP Speech Codecs			
2006 | ITU G.729.1	8 or 16	**19 … 36**	8.0 … 32.0
2009 | ITU G.719	48	**18**	32 … 128
2012 | IETF (Opus, mono/stereo)	8 - 48	**≤ 40**	8 … 128
2015 | 3GPP EVS	8 - 48	**≤ 86**	5.9 … 128

CELP: B.S. Atal, J.R. Remde | 1982
RPE-LTP: P. Vary, J. Sluyter, C. Galand | 1988

HD-Voice and the Compatibility Problem

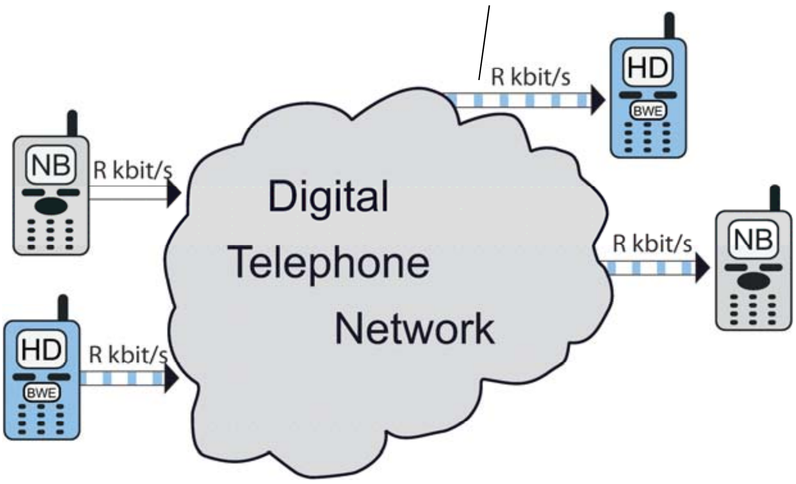
- Separate systems for NB- and HD-telephony!
- HD requires upgrading of both networks and terminals
- Long transition period with narrowband transmission



HD: Wideband device with 7.0 kHz audio quality
NB: Narrowband device with 3.4 kHz telephone quality

➢ Steganographic Side Channel

- **Hidden data transmission by watermarking**
- Bitstream, „visible“ rate R, including a „hidden“ side channel with rate S



- Hidden side channel for
  - HD-compatibility without increase of bit rate
  - frame loss concealment and/or security features
- No network upgrade

Data Hiding in CELP Codecs

- Codebook search cost function

$$\chi(\mathbf{c}) = ||\mathbf{s}'||^2 - \frac{(\mathbf{s}'^T\,\mathbf{H}\,\mathbf{c})^2}{||\mathbf{H}\,\mathbf{c}||^2}$$

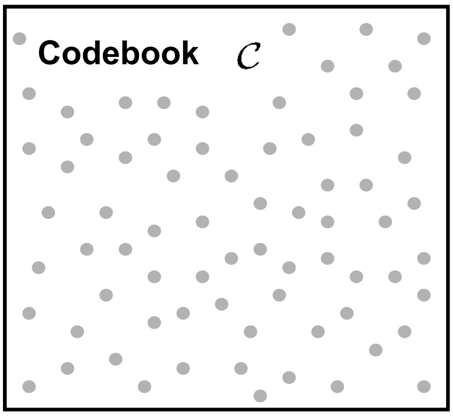
e.g. $|\mathcal{C}| = 2^{35} \approx 32 \cdot 10^9$

35 bits per 40 samples

$\mathbf{s}'$ = Target speech vector

$\mathbf{c}$ = Codebook vector

$\mathbf{H}$ = Impulse response matrix



CELP: B.S. Atal, J.R. Remde | 1982
M.R. Schroeder, B.S. Atal | 1995

Data Hiding in CELP Codecs

- Codebook search cost function

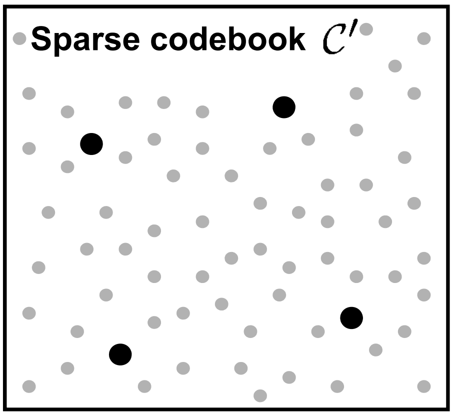
$$\chi(\mathbf{c}) = ||\mathbf{s}'||^2 - \frac{(\mathbf{s}'^T\,\mathbf{H}\,\mathbf{c})^2}{||\mathbf{H}\,\mathbf{c}||^2}$$

- **Restricted** (sparse) codebook search

$$\hat{\mathbf{c}} = \arg\min_{\mathbf{c}\in\mathcal{C}'} \chi(\mathbf{c})$$

Examined subset: $\mathcal{C}' \subset \mathcal{C}$

e.g. EFR: $|\mathcal{C}'|/|\mathcal{C}| < 10^{-6}$



Laflamme, Adoul et. al. | 1998

Data Hiding in CELP Codecs

- Codebook search cost function

$$\chi(\mathbf{c}) = ||\mathbf{s}'||^2 - \frac{(\mathbf{s}'^T\,\mathbf{H}\,\mathbf{c})^2}{||\mathbf{H}\,\mathbf{c}||^2}$$

- **Restricted** (sparse) codebook search

$$\hat{\mathbf{c}} = \arg\min_{\mathbf{c}\in\mathcal{C}'} \chi(\mathbf{c})$$

- ***Embedding of „message“ m***

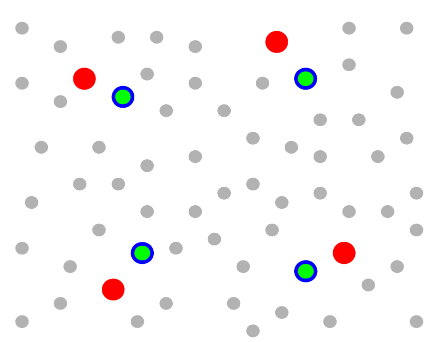
$$\hat{\mathbf{c}} = \arg\min_{\mathbf{c}\in\mathcal{C}_m} \chi(\mathbf{c})$$

$$\mathcal{C}_m \cap \mathcal{C}_{m'} = \emptyset \text{ if } m \neq m'$$

2 sub-codebooks for embedding 1 bit of message

$$|\mathcal{C}_0| = |\mathcal{C}_1| = |\mathcal{C}'|$$

Sub-codebooks, same size



- Receiver recognizes codebook, used per sub-frame

Data Hiding Applied to EFR Codec

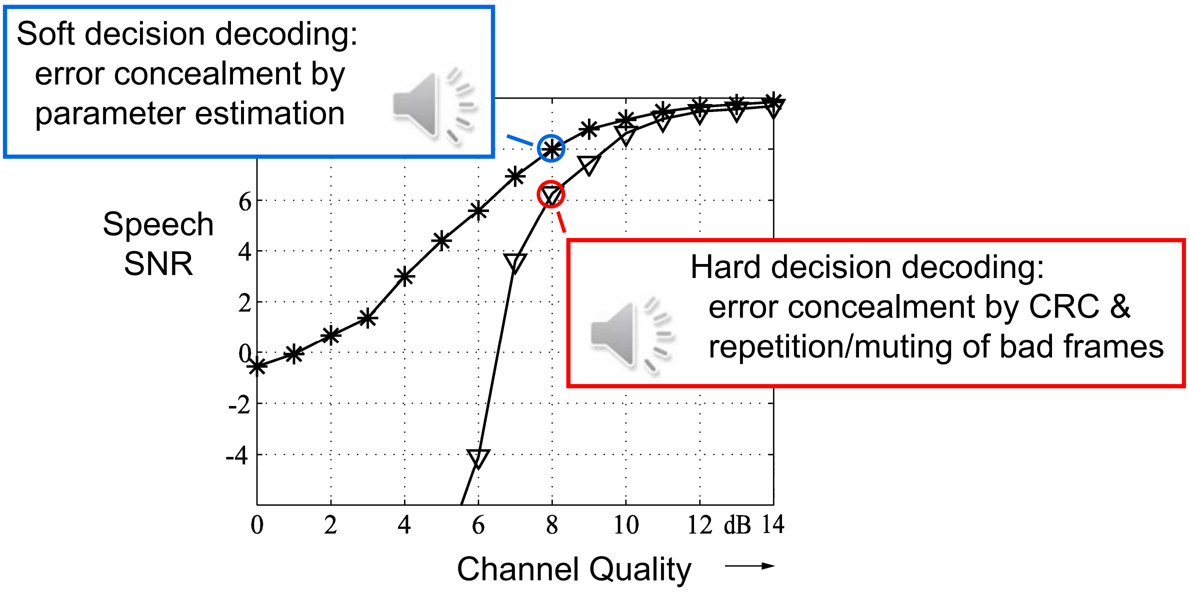
Bandwidth extension of telephone speech using hidden data channel

Example:

- Bit rate: R=12.2 kbit/s
- Compatible bit stream
- Hidden data rate:
  S=1.65 kbit/s = 8 or 9 bits/5 ms
- 2^9 different (algebraic) sub-codebooks
- Bandwidth extension by noise excitation of a synthesis filter
- No audible degradation in NB decoder

➢ Error Concealment

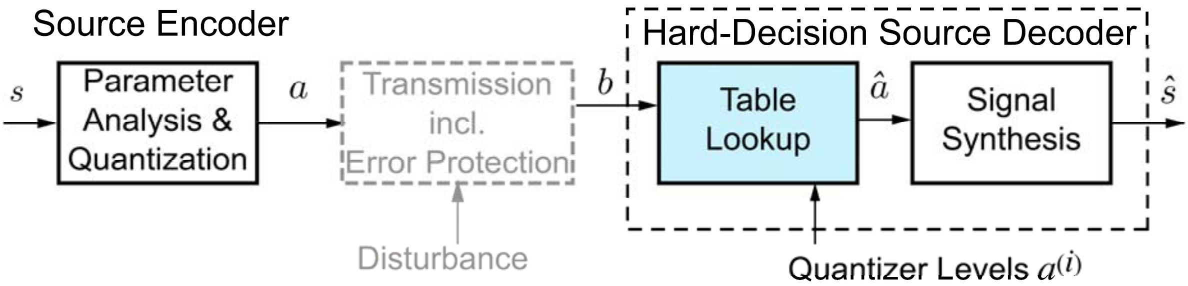
- ❑ GSM *Full Rate Codec* (13.0 kbit/s)
- ❑ GSM channel coding, modulation, equalization
- ❑ Typical urban channel (10 km/h)



Tim Fingscheidt | 1998

Speech Encoding and Hard Decision Decoding

- ❑ Speech encoding → quantized parameters
- ❑ Parameter decoding by table lookup

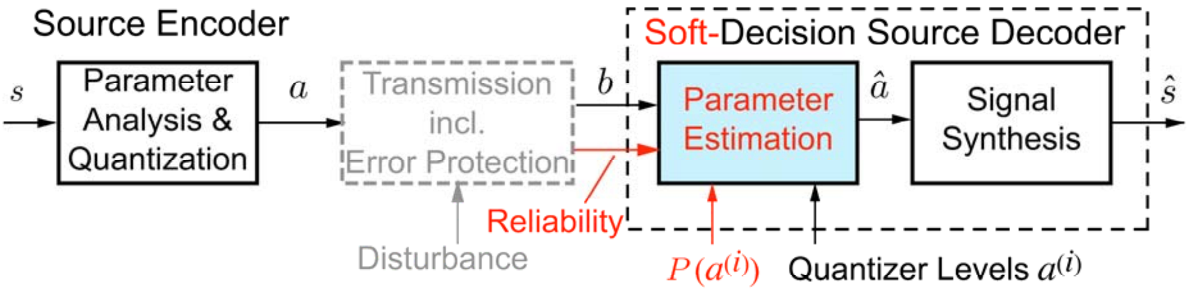


a = parameter
b = group of bits

Error Concealment by Soft Decision Decoding

- Parameter decoding by conditional estimation

$$\hat{a} = \sum_{i=1}^{2^w} a^{(i)} \cdot P(a^{(i)}|b) \qquad b = \text{group of bits}$$



s: input speech-audio signal

a: parameter, e.g. LP coefficient, gain factor, …

A priori knowledge: e.g. $P(a^{(i)})$ quantizer histogram

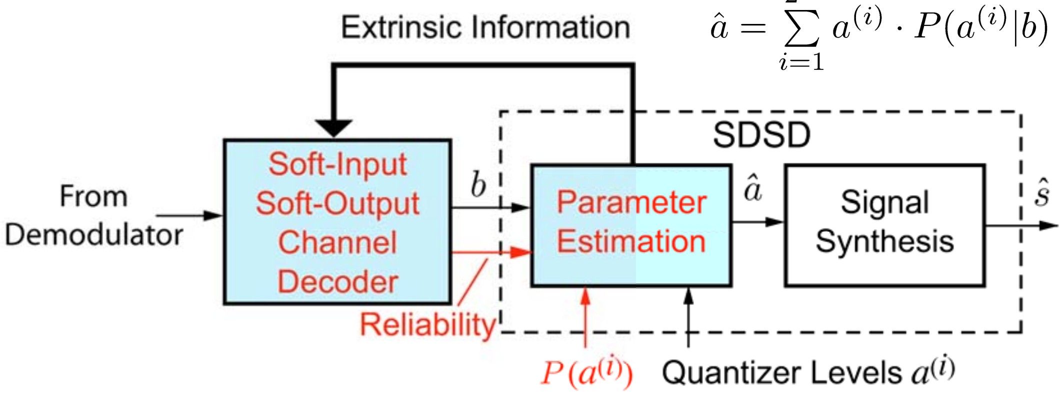
Bayes theorem: $P(a^{(i)}|b) = \frac{P(a^{(i)}) \cdot P(b|a^{(i)})}{P(b)}$

Iterative Source-Channel Decoding

Error Correction and Concealment

- Turbo processing on **bit level**

$$\hat{a} = \sum_{i=1}^{2^w} a^{(i)} \cdot P(a^{(i)}|b)$$

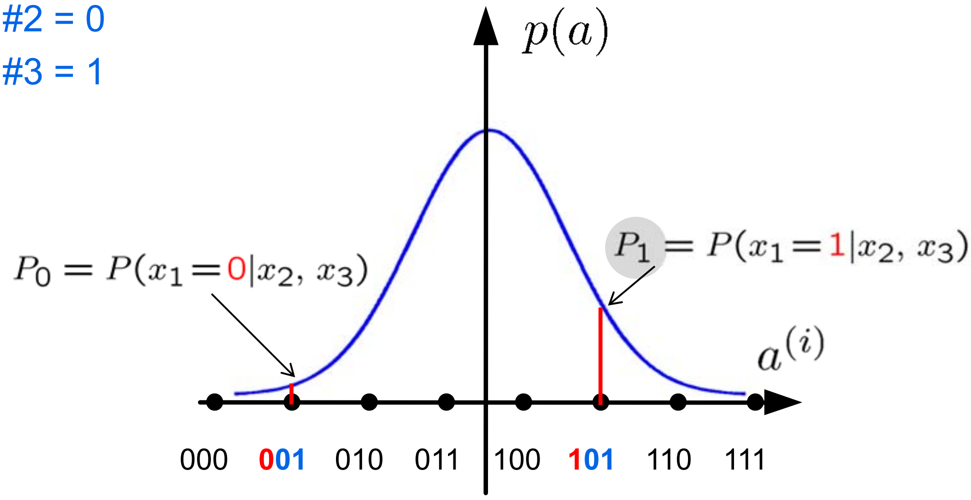


- Mean Square Estimation (MSE) on **parameter level**
- **Extrinsic information on bit level:**
  Parameter estimation supporting repeated channel decoding

Extrinsic Information from Source Decoder

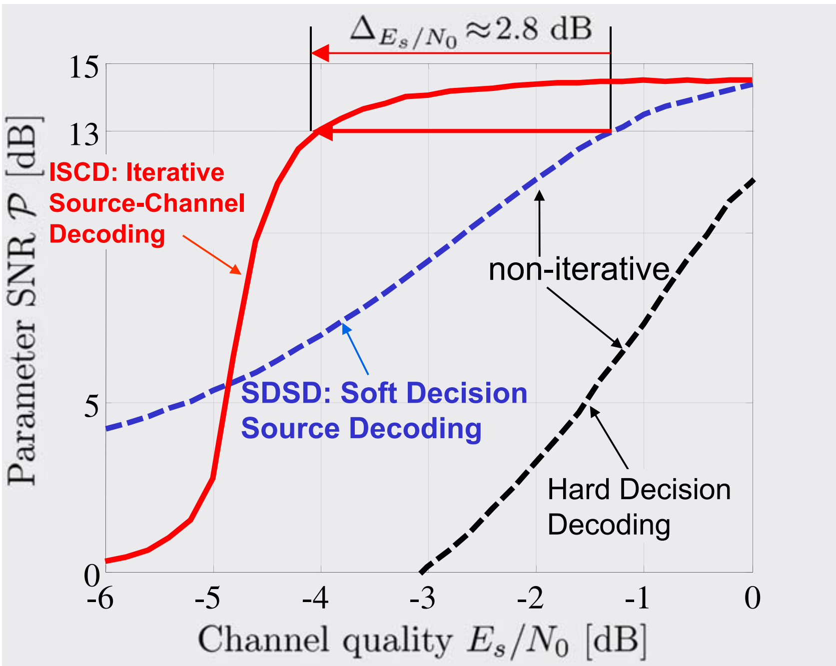
Quantization of parameter a with 8 levels / 3 bits

- Channel decoder:
  - bit #1 = ?
  - bit #2 = 0
  - bit #3 = 1



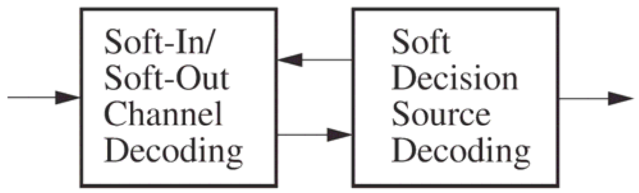
- Extrinsic information: bit #1 = 1 with probability P_1

Iterative Source-Channel Decoding (ISCD)



Example:

- A-law PCM: 8-bit per sample, 16 kHz sampling rate
- AWGN: bit error rate = 5.5 %
- Soft decision source decoding exploiting unequal parameter distribution



SDSD

+ 1 iteration

+ 2 iterations

+ 3 iterations

Laurent Schmalen | 2009

Time Warp | 1985 – 2015

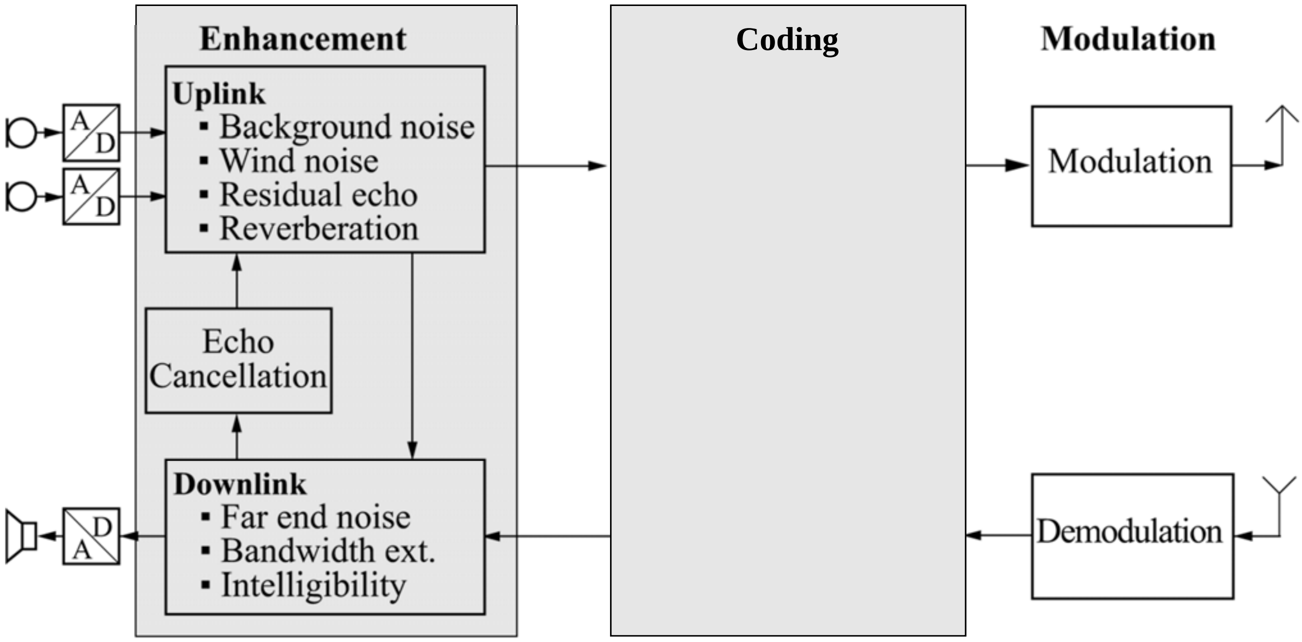
Coding

Enhancement

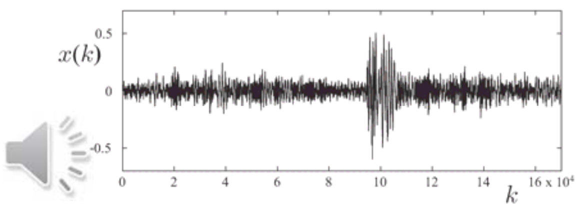
- Noise Reduction
- Acoustic Echo Control
- Intelligibility Enhancement
- Bandwidth Extension (BWE)
- Wind Noise Reduction
- Dereverberation

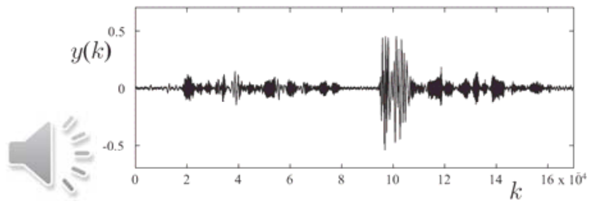
Trends

Uplink & Downlink Enhancement in a Mobile Phone

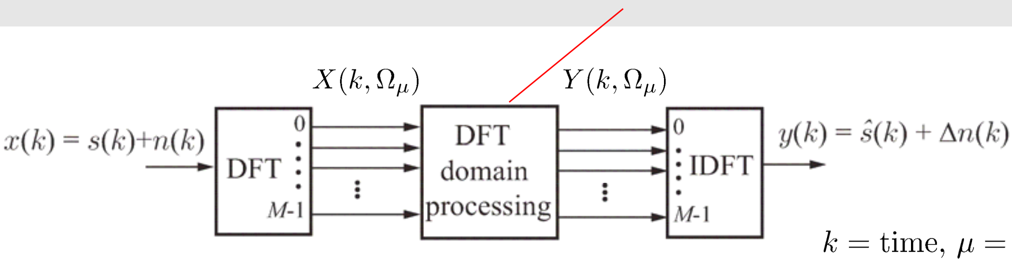


➢ Uplink Single Microphone Noise Reduction





$$Y(k,\Omega_\mu) = G_\mu(k) \cdot X(k,\Omega_\mu); \quad 0 \le G_\mu \le 1$$



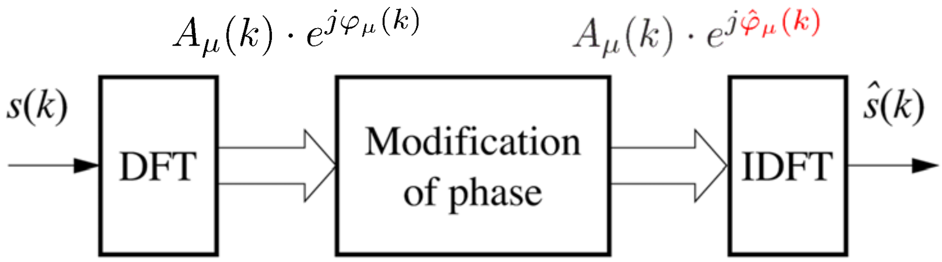
k = time, μ = frequency

- ❑ Modification of magnitude only
- ❑ Noisy phase is kept

e.g. S.F. Boll | 1979
Y. Ephraim, D. Malah | 1984/85
R. Martin | 2002
P. Wolfe, S. Godsill | 2003
T. Lotter | 2005

Relevance of Phase

- DFT length, $M = 256$, Hamming-window, overlap $M/2$
- Frame length $32\,\mathrm{ms}$



$k = \text{time},\ \mu = \text{frequency}$

Phase:

original	$\hat{\varphi}_\mu(k) = \varphi_\mu(k)$	
zero	$\hat{\varphi}_\mu(k) = 0$	
random (uniform)	$\hat{\varphi}_\mu(k) = n_0(k)$	; $-\pi \leq n_0 \leq \pi$
noisy	$\hat{\varphi}_\mu(k) = \varphi_\mu(k) + n_1(k)$	; $-\frac{\pi}{4} \leq n_1 \leq \frac{\pi}{4}$

P. Vary | 1985

Spectral Magnitude Subtraction / Weighting Rules

- Example: Wiener weights by spectral subtraction

$$Y(k,\Omega_\mu) = G_\mu(k) \cdot X(k,\Omega_\mu); \quad 0 \leq G_\mu \leq 1$$

$$G_\mu(k) = \frac{E\{|S_\mu(k)|^2\}}{E\{|S_\mu(k)|^2\} + E\{|N_\mu(k)|^2\}}$$

$$\approx \frac{E\{|X_\mu(k)|^2\} - \hat{E}\{|N_\mu(k)|^2\}}{E\{|X_\mu(k)|^2\}}$$

- Main problem: Estimation of short-term noise power spectrum

$$\hat{E}\{|N_\mu(k)|^2\}$$

$E\{|X_\mu(k)|^2\}$ = short-term expectation

More Spectral Magnitude Weighting Rules ….

- MMSE [Ephraim & Malah, 1984]

  $$G_{E\&M} = \frac{1}{\gamma} \cdot \sqrt{\nu} \cdot \Gamma(1.5) \cdot F_1(-0.5, 1, -\nu) \qquad \gamma = \text{a posteriori SNR}$$

- Log. MMSE [Ephraim & Malah, 1985]

  $$G_{E\&M} = \frac{\eta}{1+\eta} \cdot e^{\frac{1}{2}\int_{\nu}^{\infty} e^{-t}/t \, \mathrm{d}t} \qquad \eta = \text{a priori SNR} \qquad \nu = \gamma \cdot \frac{\eta}{1+\eta}$$

- MMSE with super-Gaussian models [Martin, 2002]

  $$\hat{S} = \mathrm{E}\{S|X\} = F_M(X, \sigma_N^2, \sigma_S^2)$$

- MAP with parametric PDF model [Lotter, 2003]

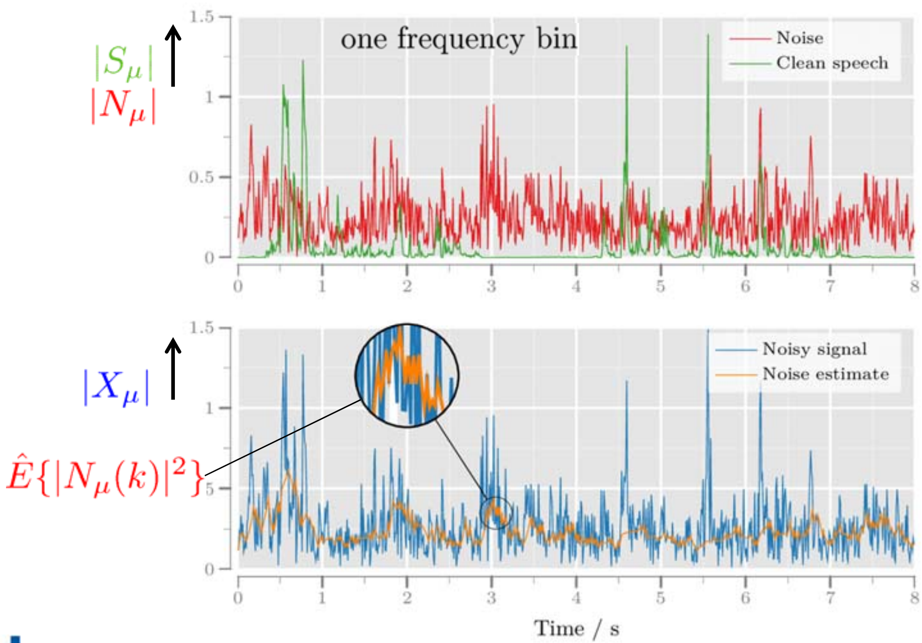
  $$G_L = u + \sqrt{u^2 + \frac{\nu - 1/2}{2\gamma}} \text{ with } u = \frac{1}{2} - \frac{\rho}{4\sqrt{\gamma \cdot \eta}} \qquad \mu, \rho = \text{parameters of PDF model}$$

- Dual Kalman filter [Esch 2012]

  $$\mathbf{K}(k) = \mathbf{P}(k)\mathbf{C}^H(k)\Big(\mathbf{C}(k)\mathbf{P}(k)\mathbf{C}^H(k) + \mathbf{\Psi}_{ss}(k)\Big)^{-1}$$

Estimation of $\hat{E}\{|N_\mu(k)|^2\}$ by "Minimum Tracking"

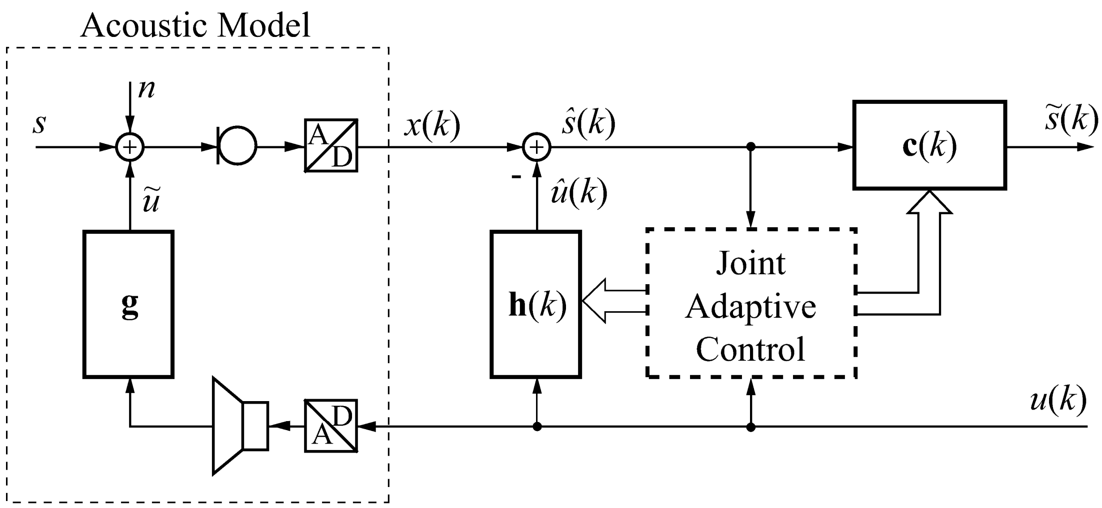
- Example: **Baseline Tracing** of slow variations [Heese, 2015]
  - Performing like a delta modulator in the log. amplitude domain
  - Low complexity implementation in the linear amplitude domain



Minimum Tracking:
Wolfgang Brox | 1983
Gerhard Doblinger | 1995
Rainer Martin | 2001
Timo Gerkmann | 2012
Florian Heese | 2015

➢ Uplink Joint Acoustic Echo & Noise Control

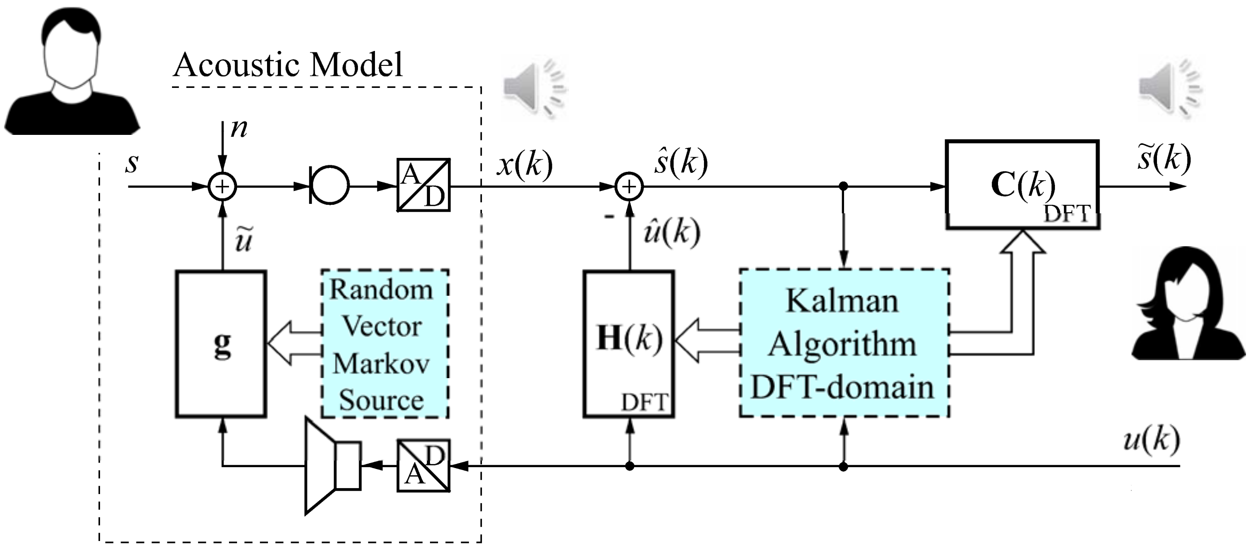
- Acoustic path $\mathbf{g}$
- Echo canceller $\mathbf{h}(k)$
- Auxiliary postfilter $\mathbf{c}(k)$ – reduction of residual echo *and* noise
- Joint adaptive control



R. Martin | 1994
F. Capman, J. Boudy, P. Lockwood | 1996

Kalman Filter Approach to Acoustic Echo Control

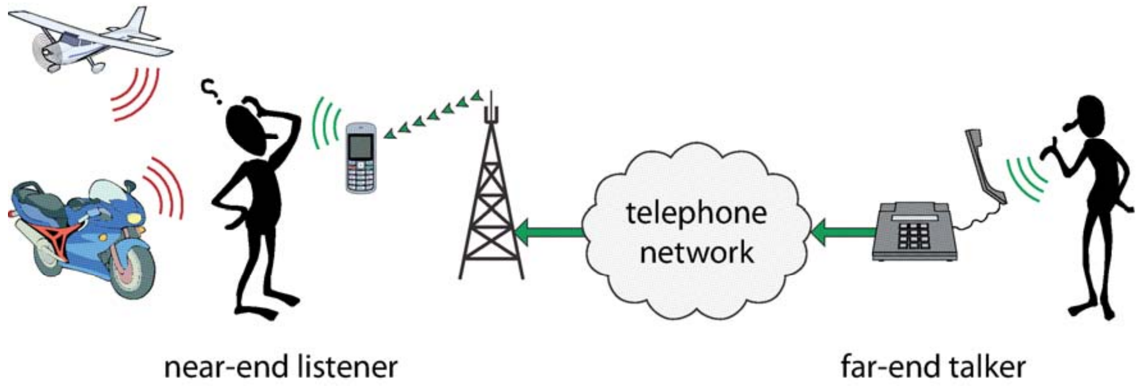
- Room impulse response as a random process
- Far end speech as a deterministic input
- DFT Domain implementation



Gerald Enzner | 2006

➢ Downlink Intelligibility Enhancement

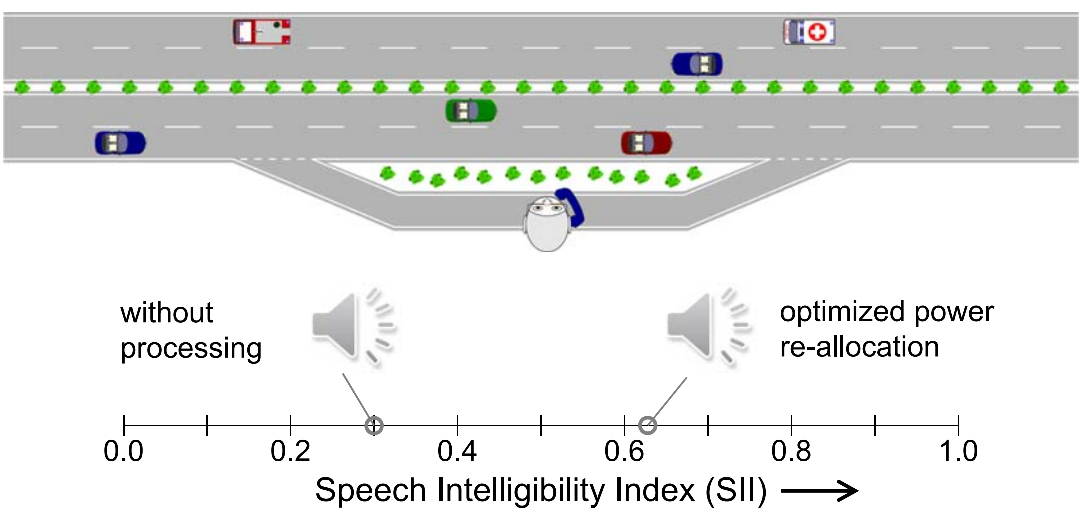
- *Near end* listener experiences reduced speech intelligibility



- **Problem:** Clear far-end speech less intelligible in near-end noise
- **Solution:** Adaptive, *frequency selective* speech amplification depending on background noise
- **Optimization criterion: Speech Intelligibility Index (SII)**

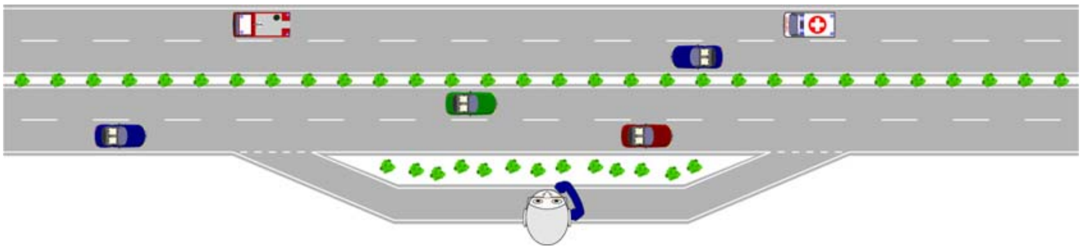
Near-End Listening Enhancement (NELE)

- Spectral power re-allocation exploiting psychoacoustics
- Optimization constraints: power limitation (ear and loudspeaker)



Near-End Listening Enhancement (NELE)

- Spectral power re-allocation exploiting psychoacoustics
- Optimization constraints: power limitation (ear and loudspeaker)



- No increase of the total audio power
- Intelligibility (modified rhyme test by Sotschek)

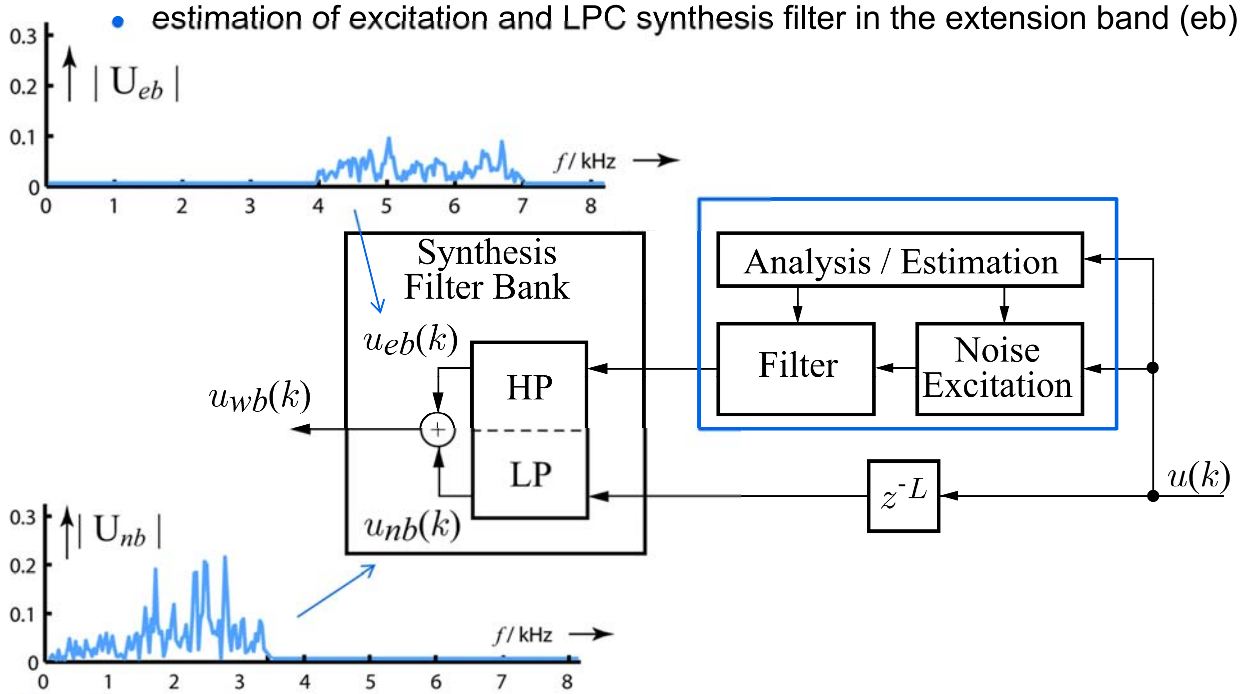
	NELE off	NELE on
Intelligibility	**29.8%**	**67.2%**

- Significant reduction of listening effort

Bastian Sauert | 2006
Markus Niermann | 2015

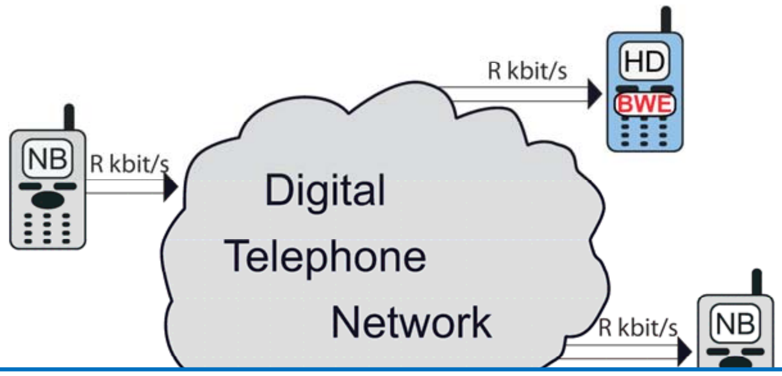
➢ Downlink Bandwidth Extension without Side-Info

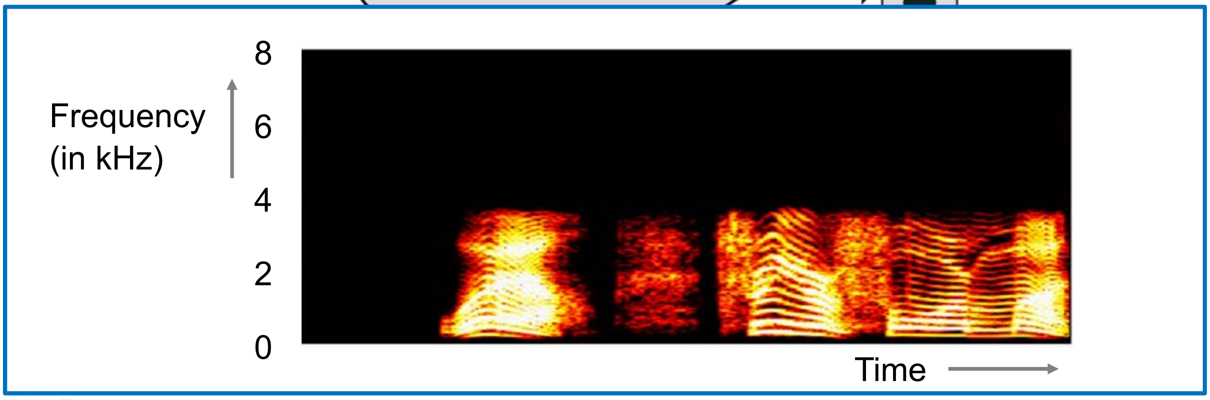
- Source-filter model for the extension band
  - analysis of the narrowband signal (300 – 3400 Hz) (nb)
  - estimation of excitation and LPC synthesis filter in the extension band (eb)



Example: BWE without Side Information

- Bandwidth extension (BWE) bridges the gap between NB and HD

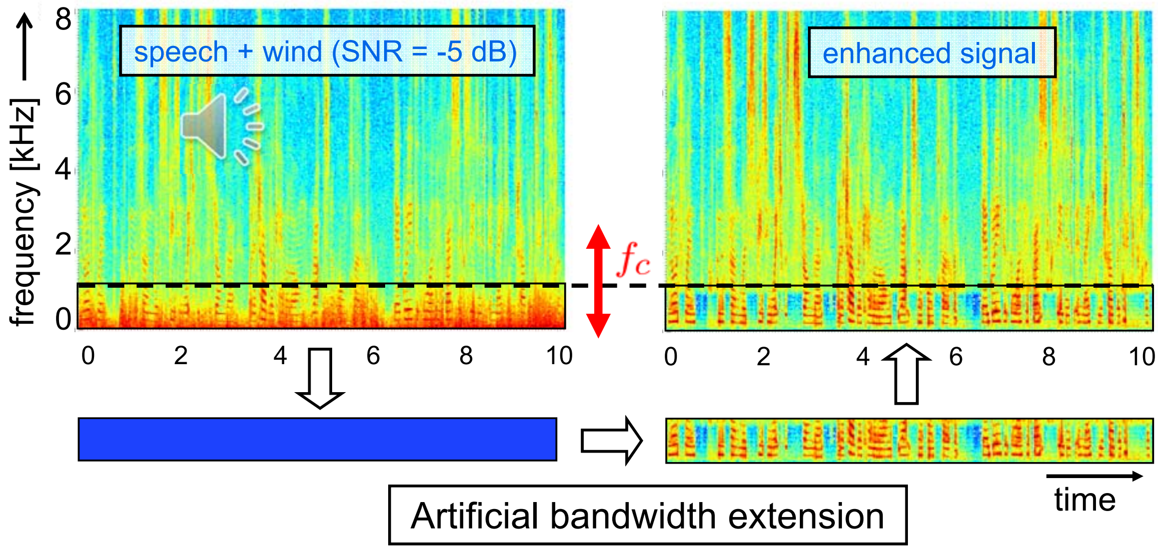




Peter Jax | 2004

➢ Uplink Wind Noise Reduction

- Wind noise = low frequency noise with $f < f_c$ (adaptive)
- Substitution of disturbed frequency band using BWE



→ Christoph Nelke, EUSIPCO 2015, Session SLP-L1: Speech Enhancement

Christoph Nelke | 2013
& EUSIPCO | 2015

Time Warp | 1985 – 2015

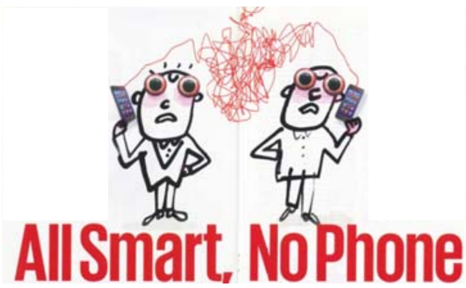
Coding

Enhancement

Trends

- Coding for Wireless Communications
- Speech & Audio Enhancement
- Applications

Trends | Coding for Wireless Communications

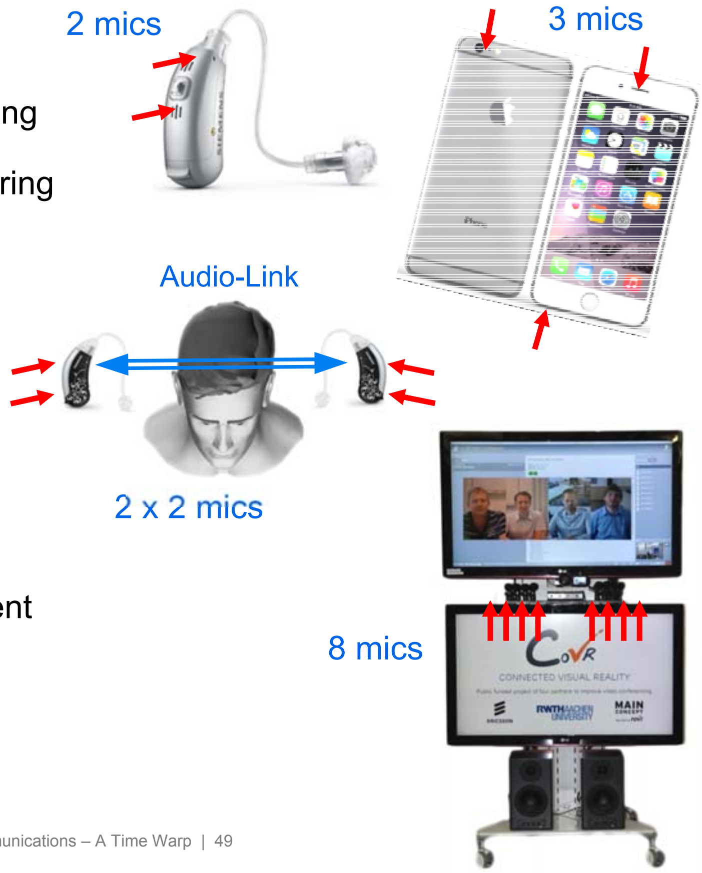


- Users rely exclusively on mobile phones
  - voice quality still an issue
- Lost focus on smartphones being also telephones
- Coding standards for wireless
  - wideband (HD) and super-wideband (HD+)
  - dual- and multi channel spatial audio codecs
- Wireless transmission goes “all IP”
  - VoLTE: voice over LTE and 5G
  - HD-voice launched / announced by 132 mobile operators
  - IP transmission eases new codecs

Jeff Hecht | IEEE Spectrum 10-2014
GSA | 04/2015

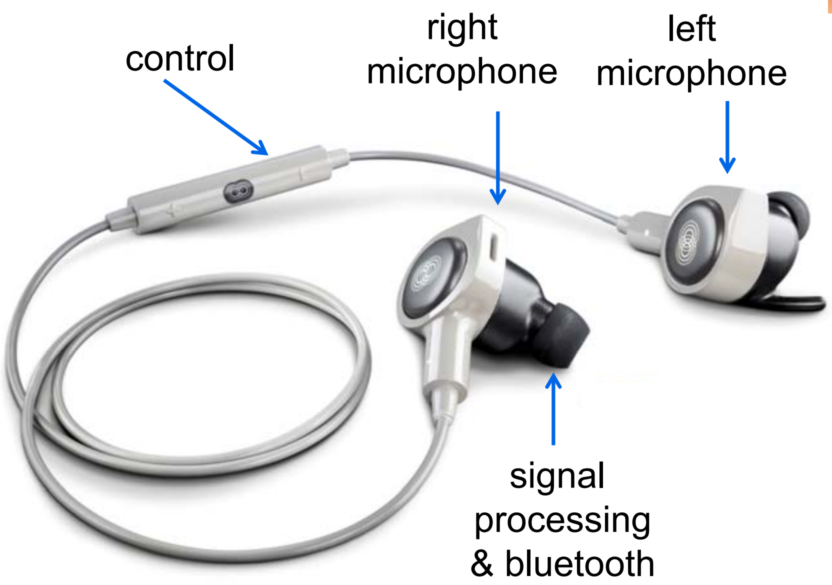
Trends | Speech & Audio Enhancement

- Dual microphone processing
- Multi microphone array processing
- Distributed wireless audio capturing
- Source separation
- Non-linear processing
- Binaural processing
- Multi-channel audio coding
- Active noise control
- Modelling of acoustic environment
- Robust speech recognition
- …



Trends | Applications

- Binaural telephony



enthusiastic user

binauric | 2016
https://www.binauric.com

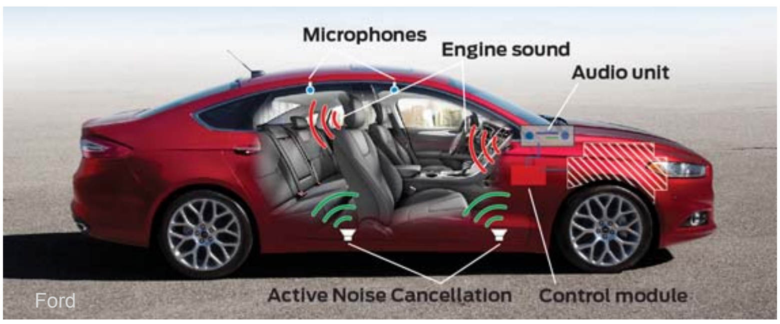
Trends | Applications

Immersive Audio / Multichannel Coding & Processing

Smart Home with Speech & Audio Components

Trends | Applications

Speech Reinforcement in Public Address Systems (NELE Approach)



In-Car Communications / Active Noise Cancellation


Wireless Speech and Audio Communications
A Time Warp
Thanks for contributions:
Marc Adrat
Christiane Antweiler
Gerald Enzner
Tim Fingscheidt
Bernd Geiser
Florian Heese
Peter Jax
Thomas Lotter
Rainer Martin
Christoph Nelke
Markus Niermann
Bastian Sauert
Magnus Schäfer
Laurent Schmalen