Deutsche Telekom Laboratories

Dimension Analysis of Wideband-transmitted Speech

Marcel Wältermann, Alexander Raake, Sebastian Möller

Deutsche Telekom Laboratories, TU Berlin, Germany

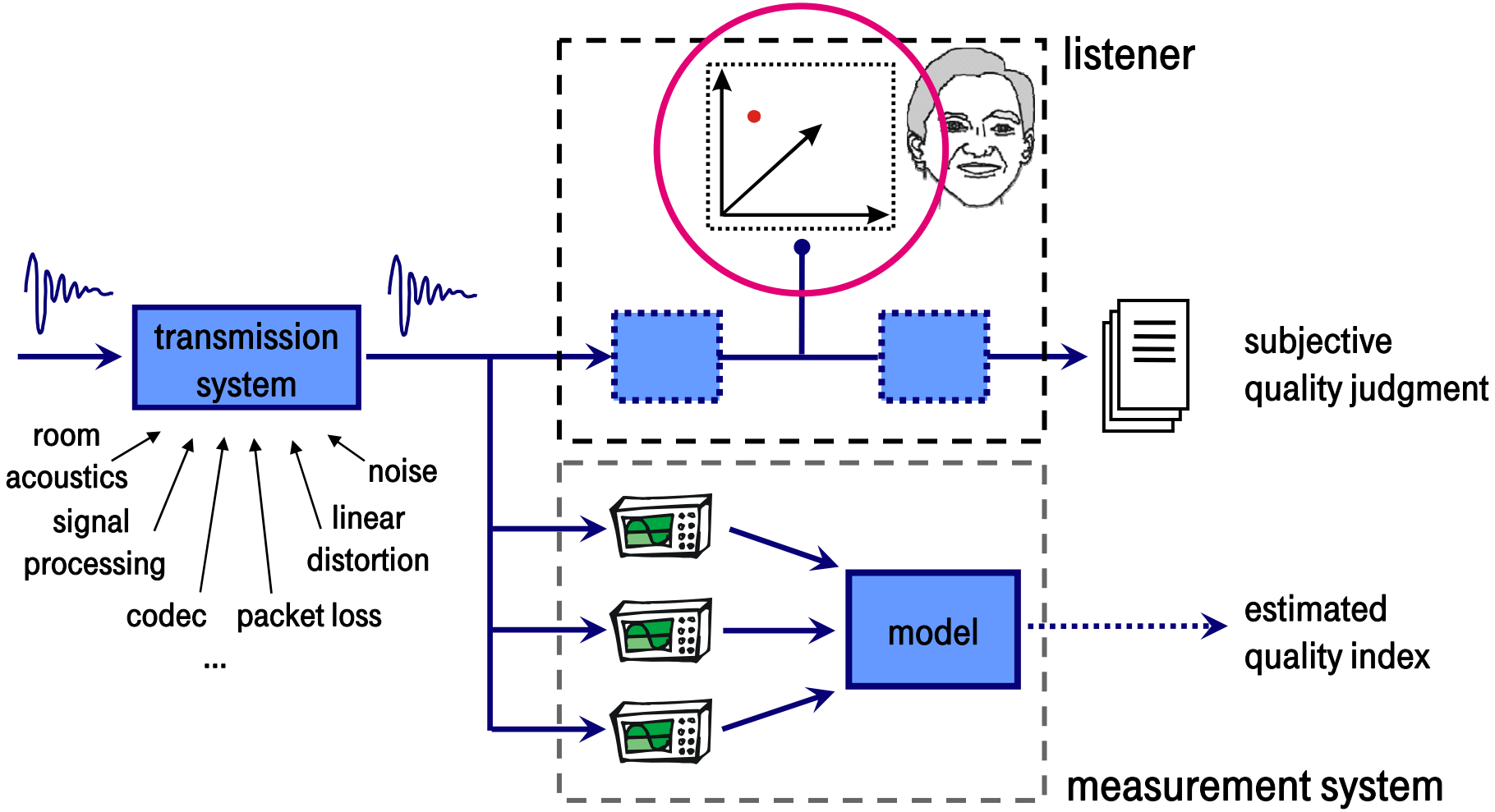
Overview

- Measuring Speech Quality: Perceptual Approach
- Quality Dimensions of Wideband-transmitted Speech
  - Experiment 1: Multidimensional Scaling (MDS)
  - Experiment 2: Semantic Differential (SD)
  - Results
- Modeling Overall Quality
- Perceptually Motivated Degradation Indicators in P.OLQA
- Example: Frequency-related Degradation Indicator
- Summary and Outlook

Overview

- Measuring Speech Quality: Perceptual Approach
- Quality Dimensions of Wideband-transmitted Speech
  - Experiment 1: Multidimensional Scaling (MDS)
  - Experiment 2: Semantic Differential (SD)
  - Results
- Modeling Overall Quality
- Perceptually Motivated Degradation Indicators in P.OLQA
- Example: Frequency-related Degradation Indicator
- Summary and Outlook

Measuring Speech Quality
Perceptual Approach



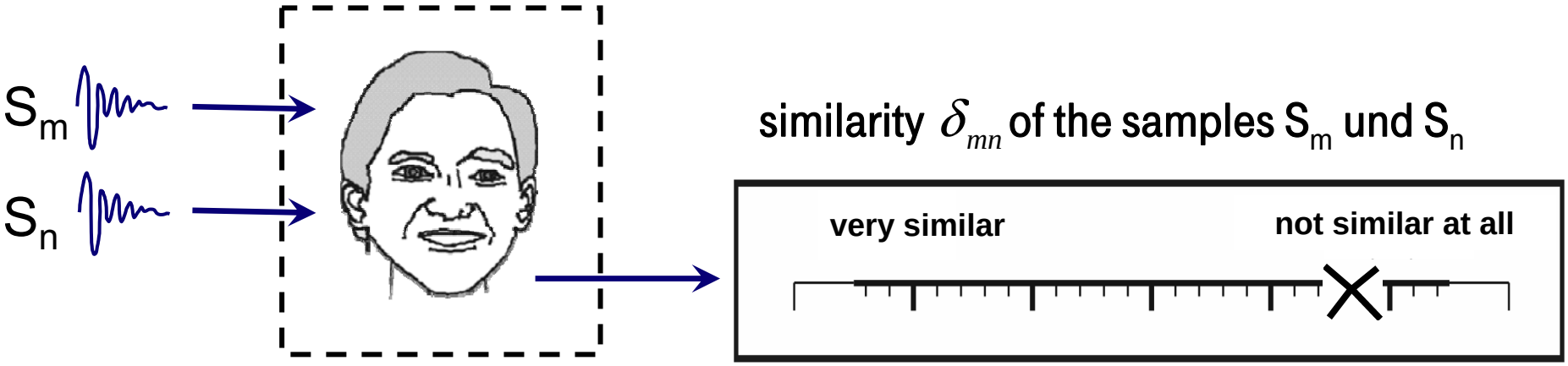
[Heute et al., 2005]

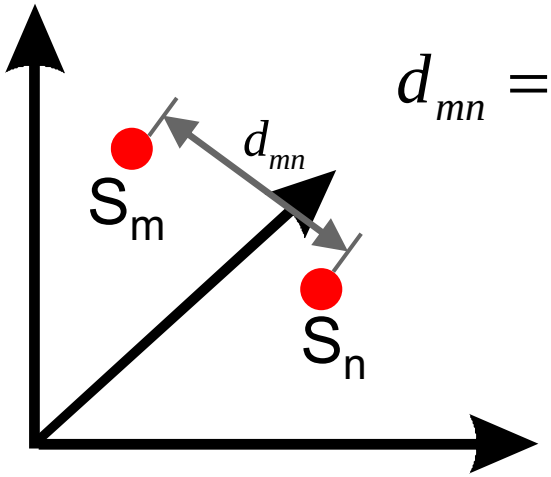
Overview

- Measuring Speech Quality: Perceptual Approach
- Quality Dimensions of Wideband-transmitted Speech
  - Experiment 1: Multidimensional Scaling (MDS)
  - Experiment 2: Semantic Differential (SD)
  - Results
- Modeling Overall Quality
- Perceptually Motivated Degradation Indicators in P.OLQA
- Example: Frequency-related Degradation Indicator
- Summary and Outlook

Experiment 1: Multidimensional Scaling (MDS)

Principle





$$d_{mn} = f(\delta_{mn})$$

determine dimensionality, so that

- $\sum_{mn,m\neq n}(\delta_{mn} - d_{mn})^2 \rightarrow \min$

- dimensions interpretable

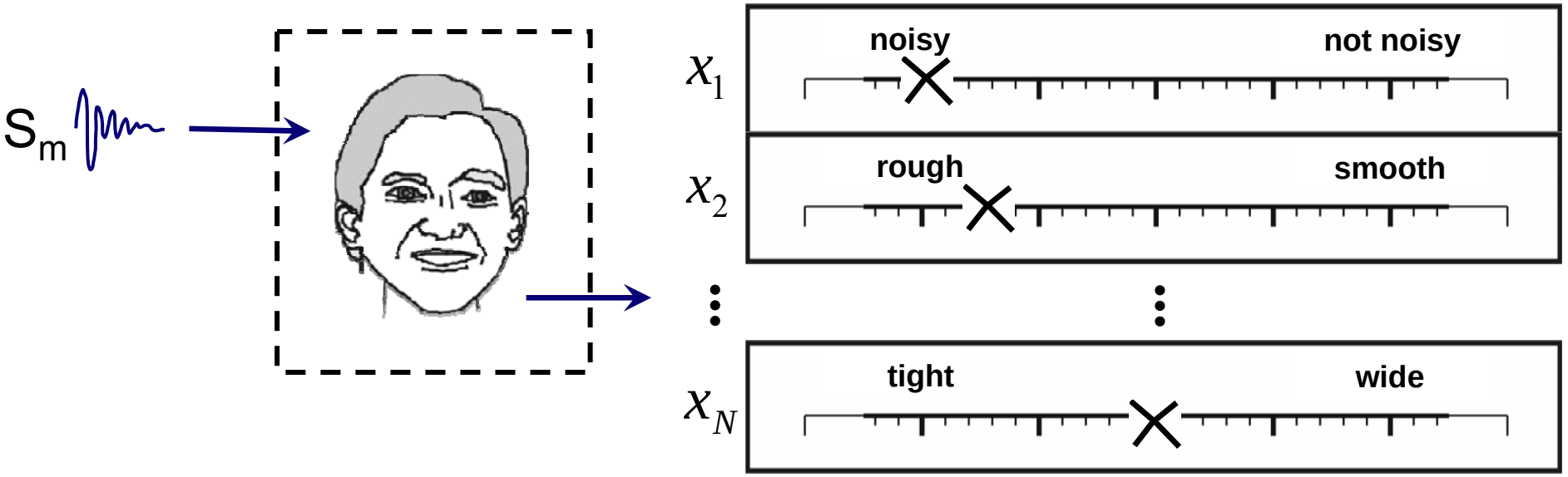
[Borg/Groenen, 2005]

Experiment 1: Multidimensional Scaling (MDS)
Details

- $I \cdot (I-1)$ pairs have to be judged, where I is the number of stimuli
- 2 speakers (male/female), $I = 14$, resulting in 364 judgments
- 19 participants (9 f, 10 m)
- INdividual Differences SCALing (rotationally invariant configuration)

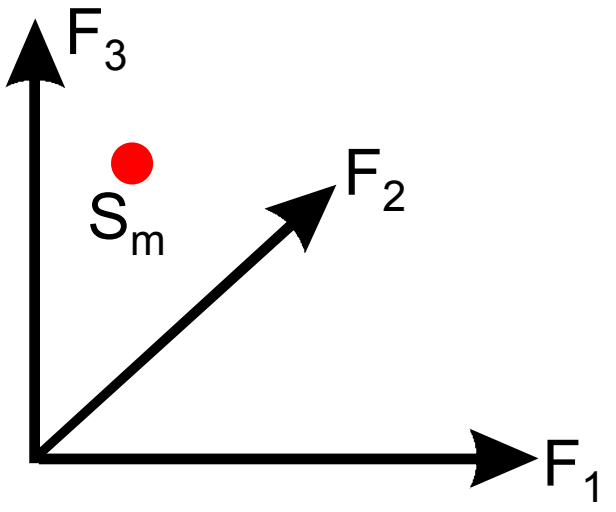
Experiment 2: Semantic Differential (SD)

Principle



[Osgood et al., 1957]

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix} \xrightarrow[M < N]{\text{factor analysis}} \begin{pmatrix} F_1 \\ F_2 \\ \vdots \\ F_M \end{pmatrix}$$



Experiment 2: Semantic Differential (SD)
Details

- 2 pre-tests with 10 “experts“ (5 f, 5 m)
  - pre-test 1: Collection of descriptive terms
  - pre-test 2: Selection of perceptively salient antonyms out of a set of the most frequently named terms in pre-test 1
  - result: 28 Antonym-pairs for semantic differential
- 28 participants (13 f, 15 m), both “experts“ and naïve listeners
- 28 · *l* judgments, where *l* is the number of stimuli
- 2 speakers (male/female), *l*=14, resulting in 784 judgments
- no individual differences were taken into account
- PCA and VARIMAX rotation

Speech Samples

Abbreviation	WB/NB	Processing elements
CLEAN	WB	Direct channel
G7221	WB	G.722.1 @ 24 kbps
AMRWB	WB	AMR-WB @ 6.6 kbps
G711	NB	G.711
BP_N	NB	G.711, 0.5 – 2 kHz bandpass
BP_B	WB	0.1 – 5 kHz bandpass
HFT_NB	NB	Hands-free terminal
HFT_WB	WB	Hands-free terminal
NC	NB	G.711, additional circuit noise
HFT_WB_N	WB	Hands-free terminal, background noise
HFT_WB_NR	WB	Hands-free terminal, noise suppression
PL20_NB	NB	G.729A, 20% packet loss
PL20_WB	WB	AMR-WB @ 23.05 kbps, 20% packet loss
ABE	WB	G.711, artificial bandwidth enhancement

Pre-analysis of the MDS and SD data

General considerations

- between-subject factor *subject-group* is statistically not significant
- interpretation of male and female speaker solution is the same

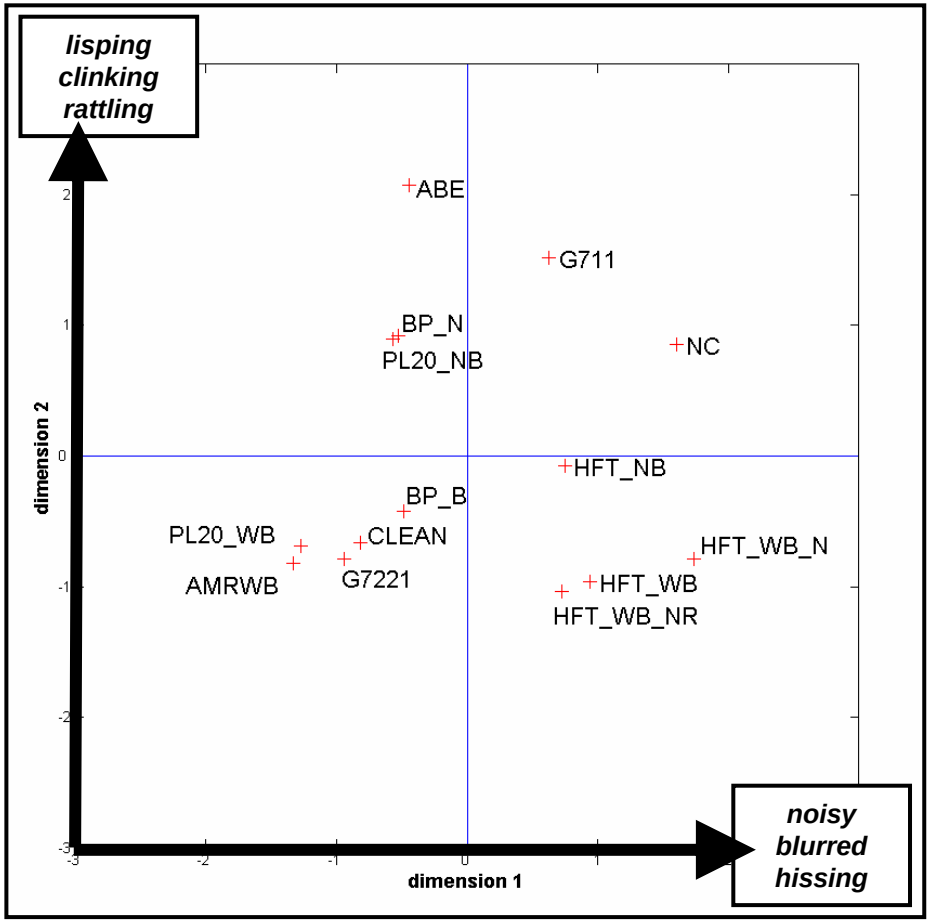
Dimensionality

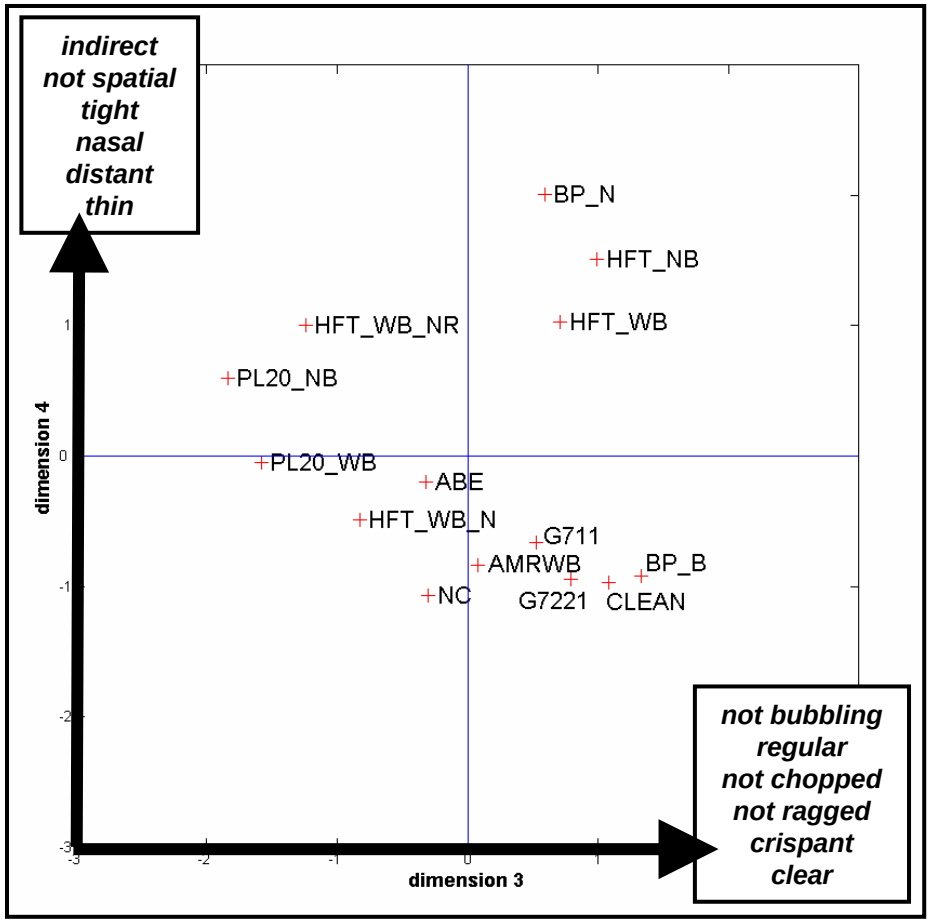
- MDS: 4-dimensional solution well interpretable (Stress = 0.19)
- SD: Kaiser criterion (eigenvalues of the correlation matrix >1) supports a 4-dimensional solution ($R^2 \approx 93\%$)

Comparison between SD and MDS solution

- mappings of the perceptual space highly resemble each other
- high correlations between single factors of SD data and dimensions of MDS data

Mapping of the Perceptual Space





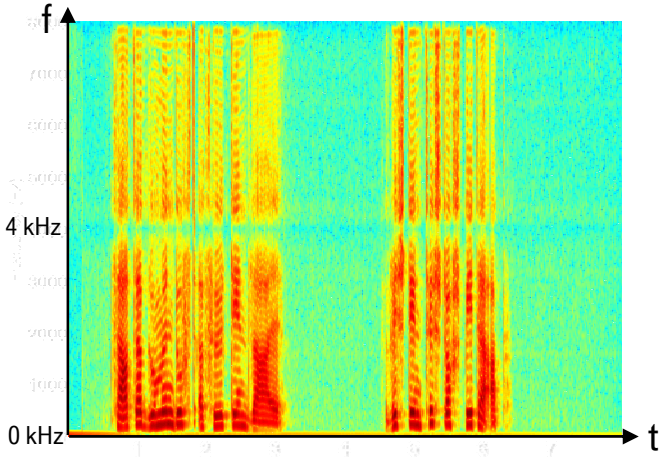
d_1: noisiness
d_2: frequency content/lisping

d_3: continuity
d_4: (in-)directness/distance

The Perceptual Attribute “Lisping”

“Continuity“, “noisiness“ and “directness“ have comparable counterparts in the narrowband-only case [Wältermann et al., 2006] however, “lisping“ does not!

“Lisping“ is the disability to pronounce sibilants properly, and instead replace them with interdentals



narrowband

artificial bandwidth enhancement

wideband

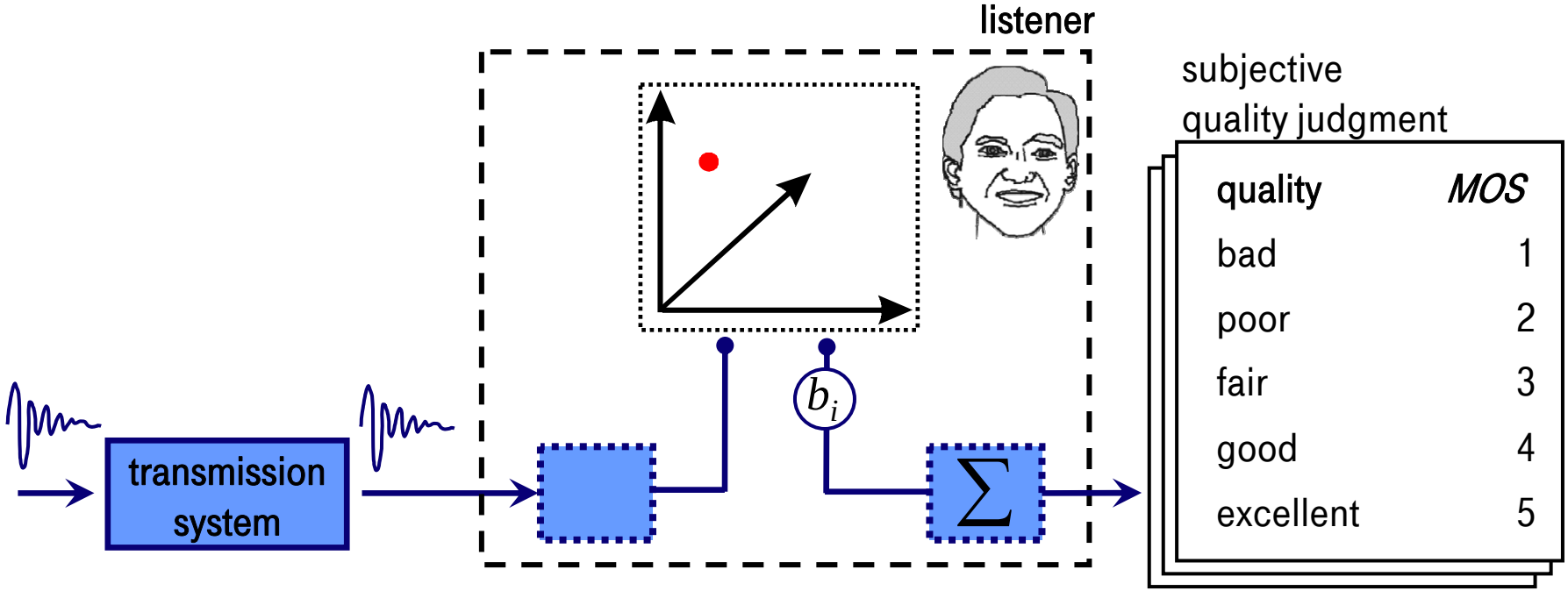
“Lisping“ can be interpreted as an **anomaly** or **lack** of high frequency components which are necessary for rendering sibilants correctly (more general label: “frequency content“)

Overview

- Measuring Speech Quality: Perceptual Approach
- Quality Dimensions of Wideband-transmitted Speech
  - Experiment 1: Multidimensional Scaling (MDS)
  - Experiment 2: Semantic Differential (SD)
  - Results
- **Modeling Overall Quality**
- Perceptually Motivated Degradation Indicators in P.OLQA
- Example: Frequency-related Degradation Indicator
- Summary and Outlook

Modeling Overall Quality

Further experiment: Collecting overall quality judgments
Mapping of the dimensions onto Mean Opinion Scores (MOS)



dimension	b_i
noisiness	-0.13
frequency content/lisping	-0.14
continuity	0.78
(in-)directness/distance	-0.30

$$MOS = \sum_{i=1}^{4} b_i \cdot d_i$$

$R^2 \approx 75\%$

Overview

- Measuring Speech Quality: Perceptual Approach
- Quality Dimensions of Wideband-transmitted Speech
  - Experiment 1: Multidimensional Scaling (MDS)
  - Experiment 2: Semantic Differential (SD)
  - Results
- Modeling Overall Quality
- **Perceptually Motivated Degradation Indicators in P.OLQA**
- Example: Frequency-related Degradation Indicator
- Summary and Outlook

Perceptually Motivated Degradation Indicators in P.OLQA

P.OLQA (Objective Listening Quality Assessment): Future objective quality measure standardized by ITU-T [ITU-T TD 12-57]

Optional feature: degradation indicators [ITU-T COM 12-4, 12-26, 12-53]

Perceptually motivated indicators provide:

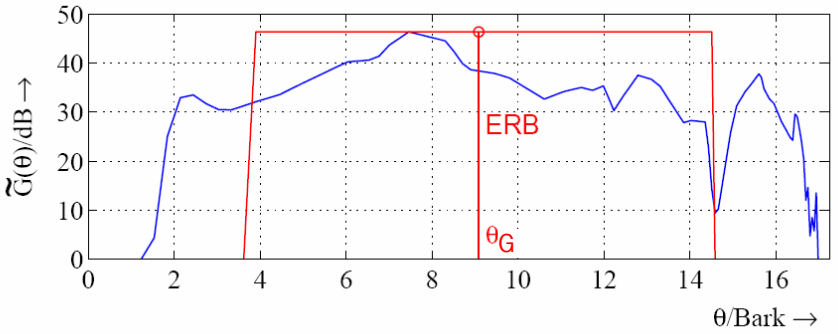
- perceptually relevant degradations
- optional output for establishing a link to physical correlates
- proven perceptual orthogonality
- assignment of degradation types to indicators in the training and benchmark phase:
  - Frequency Content/Directness: Linear filters, room-acoustic effects
  - Noisiness: Additive noise, multiplicative noise
  - Continuity: Musical tones, error (packet loss) conditions

Overview

- Measuring Speech Quality: Perceptual Approach
- Quality Dimensions of Wideband-transmitted Speech
  - Experiment 1: Multidimensional Scaling (MDS)
  - Experiment 2: Semantic Differential (SD)
  - Results
- Modeling Overall Quality
- Perceptually Motivated Degradation Indicators in P.OLQA
- **Example: Frequency-related Degradation Indicator**
- Summary and Outlook

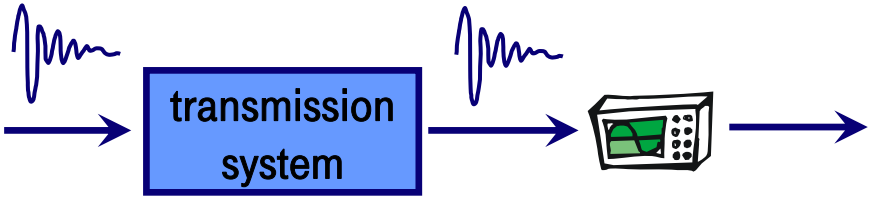
Example: Degradation Indicator "Frequency content"/"Directness"

Pilot Study: Exploit two simple physical parameters to capture the frequency-related dimensions



$$ERB = \frac{\text{area}\{\tilde{G}(\theta)\}}{\max\{\tilde{G}(\theta)\}} \qquad \theta_G = \frac{\int \tilde{G}(\theta)\cdot\theta\, d\theta}{\int \tilde{G}(\theta)\, d\theta}$$

$\tilde{G}(\theta)$ is the smoothed and bandlimited version of the gain function $G(\theta)$ of a system



$$\hat{d}_i = c_0 + c_1 \cdot ERB + c_2 \cdot \theta_G$$

$\hat{d}_2$: $|c_1| << |c_2|$

$\hat{d}_4$: $|c_1| >> |c_2|$

	d_1	d_2	d_3	d_4
r	0.64	**0.86**	0.48	**0.70**
RMSE	0.77	**0.51**	0.88	**0.71**

cf. [Scholz et al., 2005]

Overview

- Measuring Speech Quality: Perceptual Approach
- Quality Dimensions of Wideband-transmitted Speech
  - Experiment 1: Multidimensional Scaling (MDS)
  - Experiment 2: Semantic Differential (SD)
  - Results
- Modeling Overall Quality
- Perceptually Motivated Degradation Indicators in P.OLQA
- Example: Frequency-related Degradation Indicator
- Summary and Outlook

Summary and Outlook

Perceptual approach for **wideband speech quality** measurement.

For the considered set of speech files,

4 speaker-independent dimensions could be identified:

- continuity
- (in-)directness/distance
- frequency content/lisping
- noisiness

Perceptual dimensions provide a means for defining degradation indicators in standardization process of a new objective quality measure.

Example for frequency-related dimension estimator/degradation indicator

Transition from NB to WB is not necessarily enough in order to provide a better quality in telephony!

Exploratory analysis! Increase of resolution of single dimensions needed.

Thank you!

The present study was carried out at Deutsche Telekom Laboratories, TU Berlin, Germany. It was supported by the Deutsche Forschungsgemeinschaft (DFG), grant MO 1038/5-2.

References

Borg, L.; Groenen, P.: *Modern Multidimensional Scaling – Theory and Applications*. Springer Series in Statistics, New York NY, 2005.

Heute, U.; Möller, S.; Raake, A.; Scholz, K.; Wältermann, M.: Integral and Diagnostic Speech-quality Measurement: State of the Art, Problems, and New Approaches. In: *Proc. 4th European Congress on Acoustics (Forum Acusticum 2005)*, H-Budapest, 2005.

ITU-T Contribution COM 12-4: *Speech Degradation Decomposition Using a P.862 PESQ Based Approach.* Source: TNO Telecom, Netherlands (J. G. Beerends), International Telecommunication Union, CH-Geneva, 2004.

ITU-T Contribution COM 12-26: *P.OLQA Speech Quality Degradation Decomposition Benchmark Proposal.* Source: TNO Information and Communication Technology, Netherlands (J. G. Beerends, J. M. van Vugt), International Telecommunication Union, CH-Geneva, 2006.

ITU-T Contribution COM 12-53: *P.OLQA Degradation Decomposition – Perceptual Basis for Degradation Indicators*. Source: Deutsche Telekom AG (M. Wältermann, S. Möller), International Telecommunication Union, CH-Geneva, 2007.

ITU-T Temporary Document TD 12-57: *Requirement specification for P.OLQA .* Source: Rapporteur of Question 9/12, International Telecommunication Union, CH-Geneva, 2007.

Möller, S.: *Assessment and Prediction of Speech Quality in Telecommunications*. Kluwer Academic Publishers, Boston MA, 2000.

Osgood, C.; Suci, G.; Tannenbaum, P.: *The Measurement of Meaning*. University of Illinois Press, Urbana IL, 1957.

Raake, A.: *Speech Quality of VoIP – Assessment and Prediction*. Wiley, UK-Chichester, West Sussex, 2006.

Scholz, K.; Wältermann, M.; Huo, L.; Raake, A.; Möller, S.; Heute, U.: Vergleich der instrumentellen Erfassung der Qualitätsdimension 'Direktheit/Frequenzgehalt' bei Schmalband- und Breitbandsprache. In: *7. ITG-Fachtagung Sprachkommunikation*, VDE Verlag, D-Berlin, 2006.

Wältermann, M.; Scholz, K.; Raake, A.; Heute, U.; Möller, S.: Underlying Quality Dimensions of Modern Telephone Connections. In: *Proc. 9th International Conference on Spoken Language Processing (ICSLP 2006)*, Pittsburgh PA, 2006.