

# Messtechnische Erfassung der Qualitätsdimension „Direktheit/Frequenzgehalt“ zur instrumentellen Analyse und Beurteilung von Sprachqualität

Kirstin Scholz<sup>1</sup>, Marcel Wältermann<sup>2</sup>, Lu Huo<sup>1</sup>, Alexander Raake<sup>3</sup>,  
Sebastian Möller<sup>3</sup>, Ulrich Heute<sup>1</sup>

<sup>1</sup> LNS, Christian-Albrechts-Universität, Kaiserstr. 2, D-24143 Kiel, {ks,lhu,uh}@tf.uni-kiel.de

<sup>2</sup> IKA, Ruhr-Universität Bochum, D-44780 Bochum, Marcel.Waeltermann@ruhr-uni-bochum.de

<sup>3</sup> DT Laboratories, Ernst-Reuter-Platz 7, D-10587 Berlin, {Alexander.Raake,Sebastian.Moeller}@telekom.de

## Einleitung

Wie in [1] beschrieben, entwickeln wir ein *instrumentelles* Verfahren, das die Gesamt-Qualität eines Telekommunikationsnetzes sowohl *beurteilt* als auch *analysiert*. Dieses Verfahren beruht auf der Bewertung von *Qualitätsdimensionen*, die jeweils einzelne qualitätsrelevante Eigenschaften eines Sprachsignals beschreiben. Für schmalbandige Telefon-Sprachsignale (0-4kHz) wurden in [2] drei relevante Dimensionen bestimmt: „Direktheit/Frequenzgehalt“, „Kontinuität“ und „Rauschhaftigkeit“. Die Urteile über die Einzel-Dimensionen ermöglichen eine Analyse der Systemqualität. Ein Modell zur Schätzung des Mean-Opinion-Score (MOS), der die Gesamt-Systemqualität beschreibt, soll durch eine geeignete Verknüpfung dieser Dimensionen realisiert werden. Dieser Beitrag stellt Dimensionsparameter zur instrumentellen Erfassung von „Direktheit/Frequenzgehalt“ vor. Ein erster, auf zwei der Parameter beruhender Dimensionsschätzer wird vorgestellt. Außerdem wird ein Modell der untersuchten Qualitätsdimension vorgestellt, das u.a. zur Bestimmung weiterer dimensionsrelevanter Parameter genutzt werden kann.

## Dimensionsparameter

Wir vermuten, dass die Dimension „Direktheit/Frequenzgehalt“ den qualitätsrelevanten Einfluss des Frequenzgangs eines Systems auf ein von dem System verarbeitetes Sprachsignal erfasst. Zur Erfassung von „Direktheit/Frequenzgehalt“ eines verarbeiteten Signals  $y(k)$  verwenden wir daher Parameter des Betragsfrequenzgangs  $|H'(e^{j\Omega})|$  des verarbeitenden Systems:

$$|H'(e^{j\Omega})| = 20 \cdot \lg \left[ \frac{|\Phi_{xy}(e^{j\Omega})|}{|\Phi_{xx}(e^{j\Omega})|} \right]. \quad (1)$$

Hierbei gibt  $\Omega = 2\pi \frac{f}{f_A}$  die auf die Abtastfrequenz  $f_A$  normierte Frequenz  $f$  an.  $\Phi_{xx}(e^{j\Omega})$  beschreibt das Leistungsdichtespektrum des Eingangssignals  $x(k)$  und  $\Phi_{xy}(e^{j\Omega})$  das Kreuzleistungsdichtespektrum von  $x(k)$  und  $y(k)$ . Parameter des Betragsfrequenzgangs  $|H'(e^{j\Omega})|$ , deren Einfluss auf „Direktheit/Frequenzgehalt“ wir untersuchen, sind die Bandbreite, der Schwerpunkt und die Steigung von  $|H'(e^{j\Omega})|$  sowie die Tiefe und die Breite der im Frequenzgang auftretenden Täler. Die Bestimmung dieser Dimensionsparameter wird im Folgenden beschrieben. Hierbei wird lediglich das Frequenzintervall  $f = [150\text{Hz}, 3800\text{Hz}]$  betrachtet.

## Vorverarbeitung des Frequenzgangs

**Betragsfrequenzgang  $|H(\theta)|$ :** Entsprechend dem in [3] vorgestellten Verfahren zur Modellierung des Einflusses von Bandpässen auf die Sprachqualität bestimmen wir die Dimensionsparameter als Funktion der Tonheit  $\theta$ . Hierzu wird  $|H'(e^{j\Omega})|$  in  $|H(\theta)|$  transformiert, wozu folgender Zusammenhang zwischen der Frequenz  $f$  und der Tonheit  $\theta$  verwendet wird [4]:

$$\frac{\theta}{\text{Bark}} \approx 13 \arctan \left( 0.76 \frac{f}{\text{kHz}} \right) + 3.5 \arctan \left[ \left( \frac{f}{7.5\text{kHz}} \right)^2 \right]. \quad (2)$$

**Intervall relevanter Frequenzen:** Bei der Bestimmung der Dimensionsparameter wird ausschließlich das Tonheits-Intervall  $\Delta\theta = [\theta_{\min}, \theta_{\max}]$  betrachtet, dessen Intervall-Grenzen in zwei Schritten ermittelt werden:

- 1) Modifikation von  $|H(\theta)|$  entsprechend:

$$|\hat{H}(\theta)| = \max\{|H(\theta)| + ST, 0\}. \quad (3)$$

Die Wahl des Parameters  $ST$  erfolgt derart, dass ausgehend von den Dimensionsparametern von  $|\hat{H}(\theta)|$  ein optimales lineares Modell zur Dimensions-schätzung entwickelt werden kann.

- 2) Wahl der Intervall-Grenzen  $\theta_{\min}$  und  $\theta_{\max}$  gemäß:

$$|\hat{H}(\theta < \theta_{\min})| < 0.5 \cdot \max\{\hat{H}(\theta)\}, \quad (4)$$

$$|\hat{H}(\theta > \theta_{\max})| < 0.5 \cdot \max\{\hat{H}(\theta)\}. \quad (5)$$

**Zerlegung von  $|\hat{H}(\theta)|$ :**  $|\hat{H}(\theta)|$  wird gemäß

$$|\hat{H}(\theta)| = |\tilde{H}(\theta)| + \hat{H}_R(\theta) \quad (6)$$

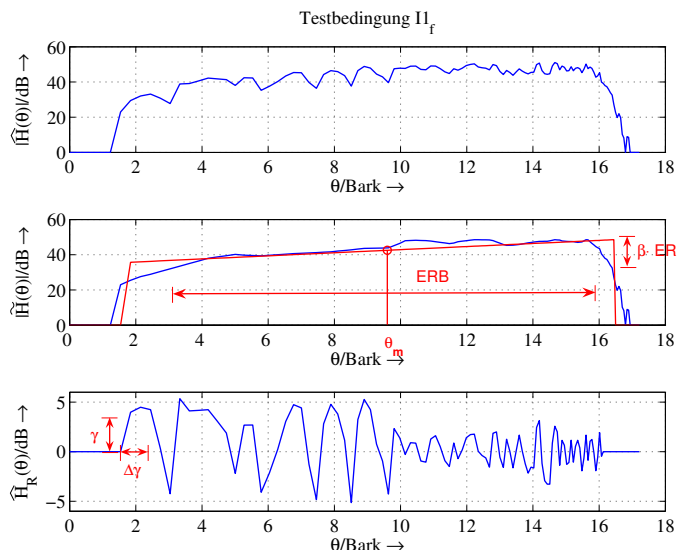
zerlegt in den geschätzten Verlauf  $\hat{H}_R(\theta)$  der Spitzen und Täler von  $|\hat{H}(\theta)|$  und einen geglätteten Verlauf von  $|\hat{H}(\theta)|$ ,  $|\tilde{H}(\theta)|$ . Ein Beispiel für eine solche Zerlegung ist in Abb. 1 dargestellt.

## Extraktion der Dimensionsparameter

Ausgehend von  $|\tilde{H}(\theta)|$  und  $\hat{H}_R(\theta)$  werden nun die Dimensionsparameter bestimmt.

**Bandbreite** Die Bandbreite von  $|\tilde{H}(\theta)|$  wird gemäß [3] mit Hilfe der äquivalenten Rechteckbandbreite (equivalent rectangular bandwidth, ERB) beschrieben:

$$ERB = \frac{\text{Fläche}\{|\tilde{H}(\theta)|\}}{\max\{|\tilde{H}(\theta)|\}}. \quad (7)$$



**Abbildung 1:** Testbedingung  $I1$  der Sprecherin aus [2]: a) Betragsfrequenzgang  $|\hat{H}(\theta)|$ , b) geglätteter Betragsfrequenzgang  $|\tilde{H}(\theta)|$  mit  $ERB$ ,  $\theta_m$  und  $\beta$  sowie c) geschätzter Verlauf  $\hat{H}_R(\theta)$  der Täler und Spitzen mit  $\gamma$  und  $\Delta\gamma$ .

**Mittentonheit** Der Schwerpunkt von  $|\tilde{H}(\theta)|$  gibt die Mittentonheit  $\theta_m$  der  $ERB$  an.

**Steigung** Die Steigung  $\beta = \text{sign}(\beta) \cdot |\beta|$  von  $|\tilde{H}(\theta)|$  wird mittels des Modells eines zu  $|\tilde{H}(\theta)|$  äquivalenten Trapezes bestimmt. Der Betrag der Steigung  $|\beta|$  ergibt sich zu:

$$|\beta| = 2 \cdot \frac{\max\{|\tilde{H}(\theta)|\} \cdot \Delta\theta - \text{Fläche}\{|\tilde{H}(\theta)|\}}{\Delta\theta^2}. \quad (8)$$

Das Vorzeichen  $\text{sign}(\beta)$  beträgt  $\text{sign}(\beta) = \pm 1$ , falls gilt  $|\theta_m - \theta_{\min}| \geq |\theta_m - \theta_{\max}|$ . Andernfalls gilt  $\text{sign}(\beta) = 0$ .

**Tiefe und Breite auftretender Täler** Anhand von  $\hat{H}_R(\theta)$  werden die Tiefe auftretender Täler,  $\gamma$ , sowie deren Breite,  $\Delta\gamma$ , bestimmt:

$$\gamma = \frac{\text{Fläche zwischen Einhüllenden der Maxima und Minima}}{\Delta\theta}, \quad (9)$$

$$\Delta\gamma = 2 \cdot \Delta\theta / (\text{Anzahl der Nulldurchgänge}). \quad (10)$$

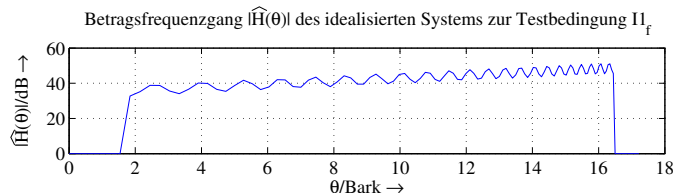
Die Dimensionsparameter des in Abb. 1a) dargestellten Frequenzgangs sind den Abb. 1b) und 1c) zu entnehmen.

## Dimensionsschätzer

Ausgehend von den 14 in [2] untersuchten Sprachproben kann ein erster Dimensionsschätzer mit *drei* Parametern entwickelt werden. Dieser Schätzer  $\widehat{DF}$  verwendet neben einer Konstanten die Dimensionsparameter  $ERB$  und  $\theta_m$  zur Schätzung von „Direktheit/Frequenzgehalt“. Für die Sprachproben der Sprecherin wurde mittels linearer Regression folgender Schätzer  $\widehat{DF}$  entwickelt:

$$\widehat{DF} = 16.1303 - 0.3516 \cdot \frac{ERB}{\text{Bark}} - 1.2705 \cdot \frac{\theta_m}{\text{Bark}}. \quad (11)$$

Für die Sprachproben des Sprechers erhalten wir eine Korrelation von  $\rho \approx 0.94$  und einen *Root Mean-Square*



**Abbildung 2:** Betragsfrequenzgang  $|\hat{H}(\theta)|$  des idealisierten Systems zur Testbedingung  $I1$  der Sprecherin aus [2].

*Error* von  $RMSE \approx 0.32$  zwischen den Ergebnissen des auditiven Tests und den Ergebnissen des Schätzers.

Zur weiteren Untersuchung des Einflusses von  $ERB$  und  $\theta_m$  sowie der bisher nicht in den Schätzer integrierten Parameter auf „Direktheit/Frequenzgehalt“ wurde ein idealisiertes System dieser Dimension entwickelt.

## Idealisiertes System

Das idealisierte System erlaubt die Vorgabe definierter Werte  $ERB$ ,  $\theta_m$ ,  $\beta$ ,  $\gamma$  und  $\Delta\gamma$ . Es ermöglicht, den Einfluss der Dimensionsparameter auf „Direktheit/Frequenzgehalt“ systematisch zu untersuchen.

Der Betragsfrequenzgang des idealisierten Systems mit den Dimensionsparametern der in Abb. 1 gezeigten Testbedingung  $I1$  der Sprecherin ist in Abb. 2 dargestellt.

## Zusammenfassung und Ausblick

In diesem Beitrag haben wir die Qualitätsdimension „Direktheit/Frequenzgehalt“ sowie zugehörige Dimensionsparameter vorgestellt. Ein erster Ansatz zur instrumentellen Schätzung dieser Dimension basierend auf  $ERB$  und  $\theta_m$  wurde präsentiert. In der nachfolgenden Arbeit wird der Einfluss der verschiedenen Dimensionsparameter auf „Direktheit/Frequenzgehalt“ mittels des idealisierten Systems systematisch untersucht werden. Ausgehend von den Ergebnissen dieser Untersuchungen soll der instrumentelle Schätzer optimiert werden.

Diese Arbeit wurde unterstützt von der Deutschen Forschungsgemeinschaft (HE 4465 und MO 1038).

## Literatur

- [1] 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. Forum Acusticum 2005, Budapest 2005, pp. 1695-1700.
- [2] Wältermann, M.; Scholz, K.; Raake, A.; Heute, U.; Möller, S.: *Qualitätsdimensionen bei der Sprachübertragung in modernen Telekommunikationsnetzen*. DAGA 2006, Braunschweig 2006.
- [3] Raake, A.: *Assessment and Parametric Modelling of Speech Quality in Voice-over-IP Networks*. Dissertation, Ruhr-Univ., Bochum 2004. Wird erscheinen als: *Perceived Quality of VoIP - Measurement and Prediction*. Chichester, West Sussex: Wiley, 2006.
- [4] Zwicker, E.; Fastl, H.: *Psychoacoustics: Facts and Models*. Springer, Berlin, 1999.