# T-Labs Series in Telecommunication Services

It is the aim of the Springer Series in Telecommunication Services to foster an interdisciplinary exchange of knowledge addressing all topics which are essential for developing high-quality and highly usable telecommunication services. This includes basic concepts of underlying technologies, distribution networks, architectures and platforms for service design, deployment and adaptation, as well as the users' perception of telecommunication services. By taking a vertical perspective over all these steps, we aim to provide the scientific bases for the development and continuous evaluation of innovative services which provide a better value for their users. In fact, the human-centric design of high-quality telecommunication services – the so called "quality engineering" – forms an essential topic of this series, as it will ultimately lead to better user experience and acceptance. The series is directed towards both scientists and practitioners from all related disciplines and industries.

** Indexing: books in this series are indexing in Scopus **

More information about this series at http://www.springer.com/series/10013

Stefan Uhrig

# Human Information Processing in Speech Quality Assessment

Springer

T LABS

Stefan Uhrig
Quality and Usability Lab
Technische Universität Berlin
Berlin, Germany

ISSN 2192-2810 ISSN 2192-2829 (electronic)
T-Labs Series in Telecommunication Services
ISBN 978-3-030-71388-1 ISBN 978-3-030-71389-8 (eBook)
https://doi.org/10.1007/978-3-030-71389-8



This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

*Dedicated to my parents and my brother*

# Preface

The present book titled *Human Information Processing in Speech Quality Assessment* covers the main outcomes of the author's doctoral research project, accomplished in the time period from 2017 to 2020. This monograph is a revised and extended version of the original dissertation approved by the doctoral committee at Technische Universität (TU) Berlin in December 2020. Two-thirds of the entire research work were initially conducted at Quality and Usability Lab of TU Berlin, the final third accrued during a one year research stay of the author at the Department of Electronic Systems of the Norwegian University of Science and Technology (NTNU) in 2019 and early 2020. Over those different project phases, the topic of human perceptual and cognitive processing of transmitted speech had to be attacked from multiple angles, thereby crossing disciplinary borders between quality engineering, acoustics, and psychology.

Previous publications in the *T-Labs Series in Telecommunication Services* have already been concerned with subjective quality assessment and prediction, as well as identifying neural correlates of speech and audio-visual quality perception. The present book contributes to this existing knowledge base in two ways: First, it closes a theoretical gap by specifying the functional structure of human information processing for listening-only test scenarios and tasks. Second, it propagates a new multi-method, "process-oriented" approach towards speech quality assessment, allowing to systematically analyze effects of varying transmission quality on specific internal processes. This approach is exemplified by three experimental studies demonstrating interactions between speech quality, stimulus context, and semantic speech content.

Besides the purpose of documenting the three studies (Chaps. 5, 6, and 7), this book opens with an introduction of fundamental concepts and methodologies (Chaps. 2 and 3), further describing a functional model of quality perception (Chap. 4) before finally attempting to theoretically integrate the studies' empirical results (Chap. 8).

The conceptual and methodological ideas together with the empirical findings elaborated in this book should specially interest researchers working in the fields of quality and audio engineering, psychoacoustics, audiology, and psychophysiology.

Berlin, Germany Stefan Uhrig
January 2021

# Acknowledgments

First of all, I am thankful to Prof. Dr.-Ing. Sebastian Möller, head of Quality and Usability (QU) Lab, Technische Universität (TU) Berlin, for his supervision, detailed feedback, and impetus that greatly shaped the research work presented in this book.

Also, I am grateful to Prof. Dr. Andrew Perkis, Department of Electronic Systems, Norwegian University of Science and Technology (NTNU), for his guidance and supervision, which provided the foundation to follow my research goals during a one year stay at NTNU from March 2019 to February 2020.

I want to give special thanks to Prof. Dr. Dawn Behne and Prof. Dr. Peter Svensson, Department of Psychology and Department of Electronic Systems, NTNU, for their trust and commitment during this one year research visit abroad.

I would like to thank Prof. Dr. Dietrich Manzey, Department of Psychology and Ergonomics, TU Berlin, for reviewing the present research work.

Many thanks go to Irene Hube-Achter, Yasmin Hillebrenner, and Tobias Jettkowski for their administrative and technical support during my years at QU Lab.

I am reminded of numerous conversations and collaborations with my colleagues at QU Lab. Thank you for the memorable times we spent together! In particular, I want to mention Gabriel Mittag, Falk Schiffner, Steven Schmidt, and Saman Zadtootaghaj, who were always on hand with help and advice.

Likewise, I want to thank my colleagues at NTNU—including members of three research groups: the Signal Processing and Acoustics groups in the Department of Electronic Systems, as well as the Speech, Cognition and Language Research (SCaLa) group in the Department of Psychology—for their valuable feedback in various colloquium talks. Especially with Asim Hameed and Marzieh Sorati, I shared many insightful discussions during my stay in Trondheim and beyond.

Last but not least, I wish to express my gratitude to both TU Berlin—namely Senta Maltschew and Evelina Skurski, from the International Scientific Cooperation Team of the International Affairs Office—and NTNU, whose strategic partnership program funded the scholarship that enabled this research work in the first place.

# Contents

# Acronyms

| | |
|---|---|
| ACR | Absolute category rating |
| AEP | Auditory evoked potential |
| AMR-WB | Adaptive multi-rate wideband |
| ANOVA | Analysis of variance |
| ANS | Autonomic nervous system |
| ASL | Active speech level |
| BAEP | Brainstem auditory evoked potential |
| BAQ | Basic audio quality |
| CAEP | Cortical auditory evoked potential |
| CCR | Comparison category rating |
| CMOS | Comparison mean opinion score |
| CNS | Central nervous system |
| Col | Coloration-impaired |
| cRT | Correct response time |
| DCR | Degradation category rating |
| Dis | Discontinuity-impaired |
| DMOS | Degradation mean opinion score |
| ECG | Electrocardiography |
| EDA | Electrodermal activity |
| EEG | Electroencephalography |
| EMG | Electromyography |
| ERP | Event-related brain potential |
| FA | Factor analysis |
| FER | Frame erasure rate |
| FIR | Finite impulse response |
| fMRI | Functional magnetic resonance imaging |
| HQ | High-quality |
| HRTF | Head-related transfer function |
| ICA | Independent component analysis |
| IFCN | International Federation of Clinical Neurophysiology |
| IP | Internet protocol |

| | |
|---|---|
| ITU | International Telecommunication Union |
| ITU-T | International Telecommunication Union—Telecommunication Standardization Sector |
| LMEM | Linear mixed-effects model |
| LPC | Late positive component |
| LPP | Late positive potential |
| LQ | Low-quality |
| MCN | Modified combinatorial nomenclature |
| MDS | Multidimensional scaling |
| MEG | Magnetoencephalography |
| MMN | Mismatch negativity |
| MNRU | Modulated noise reference unit |
| MOS | Mean opinion score |
| Noi | Noisiness-impaired |
| OLE | Overall listening experience |
| PCA | Principal component analysis |
| PS | Pairwise similarity |
| QoE | Quality of experience |
| QoS | Quality of service |
| RFE | Random frame erasure |
| RIR | Room impulse response |
| RT | Reaction time |
| S-R | Stimulus-response |
| SD | Semantic differential |
| SNR | Signal-to-noise ratio |
| TI | Talker identification |
| TTS | Text-to-speech |
| WI | Word intelligibility |
| WMA | World Medical Association |