Instrumental and Perceptual Evaluation of Hearing Devices – Methods and Applications

Christoph Völker

Disputation

Evaluation of Hearing Devices

Table of Contents

1) Introduction
2) Comparing Binaural Pre-processing Strategies
3) Modifications of the MUlti Stimulus test with Hidden Reference and Anchor (MUSHRA)
4) Unsupervised parameter optimization for hearing aid algorithms
5) Trait-based Hearing aid fitting and fine-tuning
Introduction
Facts about hearing impairments

- Fact sheet of WHO from 2015:
  360 million people worldwide have a disabling hearing loss
  (328 million adults & 32 million children)
  Over 5% of the world’s population (~ 7200 mio.)

- Hougaard & Ruf 2011:
  13.1% of German population are hearing impaired (self stated)
  [highest relative proportion of elderly people in Europe]

- Mick et al., 2014:
  Hearing impairment serious impact on the life of the affected persons (social isolation)

- Solution approach: Digital Hearing Aids
- Aim: Restauration of ability to communicate
Digital hearing aids often have **multidimensional** algorithms settings

- More than one parameter can be adjusted
- Example: Two parameters (“screws”) with 10 possible settings each → 100 possible configurations in total
Boon & Bane of Multidimensionality:
- Device can be fitted to each individual hearing loss and preference
- Device has too many setting possibilities to test with real persons
Evaluation of Hearing Devices

Two Information Sources

- **instrumental**
  - Machine (Models)

- **perceptual**
  - Humans
Evaluation of Hearing Devices
Table of Contents

1) Introduction
2) Comparing Binaural Pre-processing Strategies
3) Modifications of the MUlti Stimulus test with Hidden Reference and Anchor (MUSHRA)
4) Unsupervised parameter optimization for hearing aid algorithms
5) Trait-based Hearing aid fitting and fine-tuning
Comparing Binaural Pre-processing Strategies

In short …

- Comparison of eight signal pre-processing strategies …
- Perceptual assessment: SRTs for NH and HI
- Instrumental assessment: iSNR, STOI, PESQ, BSIM
- Analysis of correlation shows that single auditory model-based measures are not able to predict the algorithm benefit for the individual aided users/patients satisfactorily
Evaluation of Hearing Devices
Chapter I - Outcome

Evaluation of Hearing Devices

Machine (Models)

discrepancies

Humans
Evaluation of Hearing Devices

Aim of Thesis

Evaluation of Hearing Devices

Machine (Models)  discrepancies  Humans

C. Völker: Instrumental and Perceptual Evaluation of Hearing Devices
Comparing Binaural Pre-processing Strategies III: Speech Intelligibility of Normal-Hearing and Hearing-Impaired Listeners

Christoph Völker1,2, Anna Warzybok1,2, and Stephan M. A. Ernst1,2

Abstract
A comprehensive evaluation of eight signal pre-processing strategies, including directional microphones, coherence filters, single-channel noise reduction, binaural beamformers, and their combinations, was undertaken with normal-hearing (NH) and hearing-impaired (HI) listeners. Speech reception thresholds (SRTs) were measured in three noise scenarios (multitalker babble, cafeteria noise, and single competing talker). Predictions of three common instrumental measures were compared with the general perceptual benefit caused by the algorithms. The individual SRTs measured without pre-processing and individual benefits were objectively estimated using the binaural speech intelligibility model. Ten listeners with NH and 12 HI listeners participated. The participants varied in age and pure-tone threshold levels. Although HI listeners required a better signal-to-noise ratio to obtain 50% intelligibility than listeners with NH, no differences in SRT benefit from the different algorithms were found between the two groups. With the exception of single-channel noise reduction, all algorithms showed an improvement in SRT of between 2.1 dB in cafeteria noise and 4.8 dB in single competing talker condition. Model predictions with binaural speech intelligibility model explained 83% of the measured variance of the individual SRTs in the no pre-processing condition. Regarding the benefit from the algorithms, the instrumental measures were not able to predict the perceptual data in all tested noise conditions. The comparable benefit observed for both groups suggests a possible application of noise reduction schemes for listeners with different hearing status. Although the model can predict the individual SRTs without pre-processing, further development is necessary to predict the benefits obtained from the algorithms at an individual level.

Published in Trends in Hearing, Vol. 19, 2015
1) Introduction
2) Comparing Binaural Pre-processing Strategies
3) Modifications of the MUlti Stimulus test with Hidden Reference and Anchor (MUSHRA)
4) Unsupervised parameter optimization for hearing aid algorithms
5) Trait-based Hearing aid fitting and fine-tuning
Evaluation of Hearing Devices
Chapter II

Evaluation of Hearing Devices

Machine (Models) — discrepancies — Humans
Modifications of the MUSHRA listening test

Bitte bewerten Sie die Gesamtqualität

ausgezeichnet

gut

ordentlich

mäßig

schlecht

Ref. | A | B | C | D | E | F | G | H | I | J

Abbrechen | Pause | Weiter
Objectives

- Measurement practice at *Hörzentrum Oldenburg* shows that elder and technically unexperienced participants have problems with this test method concerning usability.

**Objectives of this study:**
Introduction and evaluation of two new GUls for the MUSHRA test …

- to increase accessibility and reliability
- ideally not influencing the results of ratings
Modifications of the MUSHRA listening test

MUSHRA (simple)
Modifications of the MUSHRA listening test

MUSHRA (simple)
Modifications of the MUSHRA listening test
MUSHRA (drag & drop)
Modifications of the MUSHRA listening test

MUSHRA (drag & drop)
## Modifications of the MUSHRA listening test

### Subject factors

- **Factors:**
  - age (young, old)
  - hearing (normal, impaired)
  - technical experience (high, low)

### Groups:

<table>
<thead>
<tr>
<th></th>
<th>G1</th>
<th>G2</th>
<th>G3</th>
<th>G4</th>
<th>G5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>young</td>
<td>old</td>
<td>old</td>
<td>old</td>
<td>old</td>
</tr>
<tr>
<td></td>
<td>normal hearing</td>
<td>normal hearing</td>
<td>normal hearing</td>
<td>impaired hearing</td>
<td>impaired hearing</td>
</tr>
<tr>
<td></td>
<td>tech. experienced</td>
<td>tech. experienced</td>
<td>tech. unexperienced</td>
<td>tech. experienced</td>
<td>tech. unexperienced</td>
</tr>
</tbody>
</table>
Comparison of MUSHRA Modifications by …

- Final Algorithm Ratings/Rankings
- System Usability Scale (Questionnaire)
- Preference Ranking
- Duration/Trial
- eGauge
Modifications of the MUSHRA listening test
Comparison with „eGauge“

- Standardized measure („gauge“) for the expertise („e-“) of a listener
- Values for reliability and discrimination ability for each participant
- Reliability: closeness of repeated ratings
- Discrimination: ability to perceive differences between stimuli
- Calculation done by using analysis of variance (ANOVA)
- By using permutation tests, we get a significance level for each measure
Results
Reliability & Discrimination: classic

eGauge (classic)

log(discrimination)

log(reliability)

I
II
III
IV
Results
Reliability & Discrimination: classic

- 12/50 doubtable
- ~24% doubtable data
- (surprisingly only) 3 technically unexperienced subjects (groups 3 and 5)
- 8(!) technically experienced subjects (groups 2 and 4)
- (even) 1 from control group
Results
Reliability & Discrimination: simple

- 19/50 doubtable
- ~39% doubtable data
- 9 technically unexperienced subjects (groups 3 and 5)
- 8 technically experienced subjects (groups 2 and 4)
- 2 from control group
Results
Reliability & Discrimination: drag & drop

- (only) 8/50 doubtable
- ~16% doubtable data
- 6 technically unexperienced subjects (groups 3 and 5)
- 2 technically experienced subjects (groups 2 and 4)
Modifications of the MUSHRA listening test
Differentiated choice of test method

<table>
<thead>
<tr>
<th></th>
<th>CLASSIC</th>
<th>SIMPLE</th>
<th>DRAG &amp; DROP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algorithm Ratings &amp; Rankings</td>
<td>only minor differences</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration / Trial</td>
<td></td>
<td></td>
<td>fastest</td>
</tr>
<tr>
<td>doubtful data (including G1)</td>
<td>24%</td>
<td>39%</td>
<td>16%</td>
</tr>
<tr>
<td>doubtful data of technically unexperienced (G3/G5)</td>
<td>16%</td>
<td>47%</td>
<td>32%</td>
</tr>
<tr>
<td>doubtful data of technically experienced (G2/G4)</td>
<td>40%</td>
<td>40%</td>
<td>10%</td>
</tr>
</tbody>
</table>
I stay with „old school“ MUSHRA
Modifications of the MUSHRA listening test

Conclusions

I’d rather use „drag & drop“
Evaluation of Hearing Devices
Chapter II - Outcome

Evaluation of Hearing Devices

Machine (Models)

discrepancies

Humans

Method for subjective testing
Modifications of the MUlti stimulus test with Hidden Reference and Anchor (MUSHRA) for use in audiology

Christoph Völker¹², Thomas Bisitz³, Rainer Huber³, Birger Kollmeier¹² & Stephan Ernst¹²

¹Abteilung Medizinische Physik, Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany, ²Cluster of Excellence ‘Hearing4All’, Oldenburg, Germany, and ³HörTech gGmbH, Oldenburg, Germany

Abstract

Objective: Two modifications of the standardised Multi Stimulus test with Hidden Reference and Anchor (MUSHRA), namely MUSHRA simple and MUSHRA dragdrop, were implemented and evaluated together with the original test method. The modifications were designed to maximise the accessibility of MUSHRA for elderly and technically non-experienced listeners, who constitute the typical target group in hearing aid evaluation. Design: Three MUSHRA variants were assessed based on subjective and objective measures, e.g. test-retest reliability, discrimination ability, time perception and overall preference. With each method, participants repeated the task to mix the quality of several hearing aid algorithms multiple times. Study sample: Fifty listeners grouped into five subject classes were tested, including elderly and technically non-experienced participants with normal and impaired hearing. Normal-hearing, technically experienced students served as controls. Results: Both modifications can be used to obtain comparable rating results. Both were preferred over the classical MUSHRA procedure. Technically experienced listeners performed best with the modification MUSHRA dragdrop. Conclusions: The comprehensive comparison of the MUSHRA variants demonstrates that the intuitive modification MUSHRA dragdrop can be generally recommended. However, considering e.g. specific evaluation demands, we suggest a differentiated and careful application of listening test methods.

Key Words: MUSHRA, perceptual evaluation, subjective evaluation, quality, hearing aid algorithms

Accepted by International Journal of Audiology (in press)
Evaluation of Hearing Devices

Chapter III

Machine (Models) -> discrepancies -> Humans

Method for subjective testing
Evaluation of Hearing Devices

Table of Contents

1) Introduction
2) Comparing Binaural Pre-processing Strategies
3) Modifications of the MUlti Stimulus test with Hidden Reference and Anchor (MUSHRA)
4) Unsupervised parameter optimization for hearing aid algorithms
5) Trait-based Hearing aid fitting and fine-tuning
Unsupervised parameter optimization
In short …

- Automated/Unsupervised model-based approach
- calculates percentile contours of instrumental measures within algorithm parameter space
- finds optimal parameter combinations in intersections of contours
- Here:
  - coherence-based noise reduction scheme (two parameters: alpha, tau)
  - equally weighting of four instrumental measures

Bane of multidimensional hearing aid algorithms
Unsupervised parameter optimization
Percentile Contours
Unsupervised parameter optimization

Percentile Contours
Unsupervised parameter optimization
Percentile Contours
Unsupervised parameter optimization
Intersection
Unsupervised parameter optimization
Minimal Intersection

Graph showing the relationship between tau (ms) and alpha, with different lines indicating SNR, SII, BINSII, and PSM.
Unsupervised parameter optimization
Minimal Intersection

![Graph showing the optimization process with various parameters and intersections.](image)

- SNR
- SII
- BINSII
- PSM
- Manual Set I
- Automated Set
Unsupervised parameter optimization

Results: Instrumental evaluation

• Comparison with two manual parameter sets
• Instr. enhancements over Input SNRs …
• Automated Set performs best (or equally) for SNR < 5 dB
Unsupervised parameter optimization
Results: SRT measurements

NoPre: No Pre-Processing

All settings: No SRT-benefit regarding NoPre

SRT-Ranking:
1) NoPre / Manual Set II
2) Manual Set I
3) Automated Set
Unsupervised parameter optimization

Conclusion

- Optimization method needs some more development
  - a) choice of models is crucial (SII = BINSII, global SNR, PESQ, HASQI …)
  - b) coherence scheme might not be optimal for demonstration purposes
  - c) not use „compromise setting“ (equal weighting), but weight the models differently
Evaluation of Hearing Devices
Chapter III - Outcome

Evaluation of Hearing Devices

Machine (Models)
Method for optimized settings

Discrepancies

Humans
Method for subjective testing
Evaluation of Hearing Devices
Chapter III - Outcome

create a revision
EO: Roeser, Ross
TIJA-2016-05-0152

- Minor Revision (01-Aug-2016)

Development and evaluation of an unsupervised model-based multidimensional parameter optimization for hearing aid algorithms
View Submission

1) Introduction
2) Comparing Binaural Pre-processing Strategies
3) Modifications of the MUlti Stimulus test with Hidden Reference and Anchor (MUSHRA)
4) Unsupervised parameter optimization for hearing aid algorithms
5) Trait-based Hearing aid fitting and fine-tuning
Trait-based Hearing aid fitting and fine-tuning

Introduction

- Traits?
  - „Noise haters“ or „distortion haters“ (Houben, 2013): different attitudes against noise or introduced distortions/artifacts
  - „Power Junkies“

- Hearing aid fitting compromise between aspects as speech quality and intelligibility – users have different preferences for those „trade-offs“ (Abrams et al., 2011)

- Assumption: Judgments regarding optimal algorithm settings (or different algorithms) by linear combination of individually weighted subjective traits
Trait-based Hearing aid fitting and fine-tuning

Concept: Learning phase

e.g. a prototypical „distortion hater“

trait-based: e.g. punish distortions

combined discrimination and classification task (CoDiCl)

generate stimuli with different distortion levels: same ratings?
When a sufficient agreement between objective and subjective assessment is found, i.e., models correlate with traits …

Hearing Aid Fitting: Usage of trait-based models with different weightings to differentiate between individual user preferences

„Prescription“ of not one single setting/preset, but a manageable amount of settings from a reasonable „path of probation“

User then finds preferred setting with an interactive intuitive procedure
  - „psychological ownership“ (Dillon et al., 2006)
  - „IKEA effect“ (Norton et al., 2012)
Trait-based Hearing aid fitting and fine-tuning
Exemplary application

- coherence NR scheme
- contours of two (trait-based) measures
  - noise hating: SII
  - distortion hating: PSM/PEMO-Q
- 21 different initial weightings
  - 100 / 0
  - 95 / 5
  - 90 / 10
  - ...
  - 10 / 90
  - ...

Diagram showing contours for SII and PSM/PEMO-Q measures.
Trait-based Hearing aid fitting and fine-tuning
Exemplary application
Evaluation of Hearing Devices
Chapter IV - Outcome

Evaluation of Hearing Devices

Machine (Models)
Method for optimized settings

Method for subjective testing

Humans

Method for subjective testing

discrepancies
### Evaluation of Hearing Devices

**Chapter IV - Outcome**

<table>
<thead>
<tr>
<th>STATUS</th>
<th>ID</th>
<th>TITLE</th>
<th>CREATED</th>
<th>SUBMITTED</th>
</tr>
</thead>
<tbody>
<tr>
<td>EO: Not Assigned</td>
<td>TIJA-2016-07-0221</td>
<td>Hearing aid fitting and fine-tuning based on estimated individual traits</td>
<td>12-Jun-2016</td>
<td>23-Jul-2016</td>
</tr>
<tr>
<td>Awaiting Admin Processing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*View Submission*
Acknowledgment

Co-Authors
Acknowledgment

Funding

Funded by:
- BMBF "Modellbasiertes Hörgerät" (01EZ1127D)
- DFG Exzellenzcluster “Hearing4All” (EXZ1077)
- DFG Forschergruppe (FOR1732)
Evaluation of Hearing Devices

Aim of Thesis & Outcome

Evaluation of Hearing Devices

Machine (Models)

Method for optimized settings

discrepancies

Humans

Method for subjective testing

C. Völker: Instrumental and Perceptual Evaluation of Hearing Devices