Listening Effort Scaling and Preference Rating for Hearing Aid Evaluation

Michael Schulte, Matthias Vormann, Kirsten Wagener, Michael Büchler, Norbert Dillier, Wouter Dreschler, Koen Eneman, Matthias Froehlich, Giso Grimm, Niklas Harlander, Volker Hohmann, Rolph Houben, Sofie Jansen, Arne Leijon, Anthony Lombard, Heleen Luts, Dirk Mauler, Marc Moonen, Henning Puder, Ann Spriet, Jan Wouters
Motivation / Outline

- Goal was to evaluate 5 different algorithms in background noise
- With normal hearing subjects as well as with hearing impaired subjects
- Speech recognition and hearing comfort
- Different perceptual qualities should be evaluated

→ Intelligibility
→ Listening effort
→ Preference
Speech Intelligibility

• Measured by means of sentence tests
  – German: 5 words with fixed structure
    Oldenburg Sentence Test (OLSA)
  – Dutch: everyday sentences (VU)
• Adaptive procedure, i.e. SNR was adapted
  to reach 50% intelligibility
Listening Effort Scaling / WP7 T4

Listening Effort

Max

Min

Speech Intelligibility
in % correct

100 %

0 %

Signal to Noise Ratio [dB SNR]
Listening Effort Scaling / WP7 T4

- 11 different SNRs
- Two different background noises
- 60 point scaling
- 7 categories
Listening Effort Scaling / WP7 T4

- 10 subjects
- Normal hearing
- Speech reception (OLSA)
- LES with OLSA sentences

Speech intelligibility for different background noises

Signal-to-noise ratio [dB SNR]

Speech intelligibility [%]
Listening Effort Scaling / WP7 T4

- different effort for the same intelligibility

![Graph showing listening effort scaling for normal hearing listeners with error bars for noise conditions.](image)
Listening Effort Scaling / WP7 T4

• Speech Intelligibility and Listening Effort seem to be DIFFERENT factors. This might also explain why e.g. Marzinzik & Kollmeier (1999) found no effect on intelligibility when using noise reduction algorithms but on Listening Effort.

• This is the case for normal hearing as well as hearing impaired persons.

• Different situations lead to different listening effort. It is not clear what causes the difference (spectra, modulation etc.). However, it seems that “naturalness” might be important (if you are used to the background noise).
Listening Effort Scaling

13 point scaling
7 subcategories, and always one empty button in between
Listening Effort Scaling

- Each algorithm and unprocessed were tested
- Setup: Living room, babble noise at 65 dB(A), S0N90/180/270
- Scaling at $-10$, $-5$, $0$, $+5$ and $+10$ dB SNR

$\Rightarrow$ 6 algorithms x 5 SNRs x test/retest = 60 ratings
Preference Rating

Please toggle between the Programmes:

A  B

Choose the programme you PREFER:

A  B

Instruction:
Choose the programme you prefer
(You can still switch Programmes ...)

Trial 1 of 30

Grade your preference:

B very much better than A
B much better than A
B better than A
B slightly better than A
B very slightly better than A

Instruction:
Please grade your preference
(You can still switch Programmes...)

Trial 1 of 30
Preference Rating

- Each algorithm tested against unprocessed
- Setup: Living room, babble noise at 65 dB(A), S0N90/180/270
- Rating at 0, +5 and +10 dB SNR

⇒ 5 algorithms x 3 SNRs x test/retest = 30 ratings
Preference Rating

The Linear Gaussian Model

- The LGM places the 6 algorithms on an interval scale
- Result of model is position of each algorithm on an interval scale
- Distance between algorithms is well defined and therefore useful for interpreting the data

Dahlquist/Leijon (2003)
Speech Recognition

Office room – 90/180/270 ° babble noise

SRT improvement (dB)

CH-UHZ
DE-HZO
BE-LEU
NL-AMC
Average
Listening Effort Scaling

Office room – 90/180/270°, babble noise, 0 dB SNR
Listening Effort Scaling

Mean over all subjects (N=112)
Living Room, 90°/180°/270°

-10 dB SNR -5 dB SNR 0 dB SNR 5 dB SNR 10 dB SNR

unprocessed
SC1
SC2
BSS
MWF
COH

extreme effort
much effort
considerable effort
moderate effort
little effort
very little effort
no effort
Listening Effort Scaling

Mean over all subjects (N=113)
Living Room, 90°/180°/270°

Algorithm better than unprocessed

-10 dB SNR -5 dB SNR 0 dB SNR 5 dB SNR 10 dB SNR

Unprocessed better
Preference Rating

Office room – 90/180/270° babble noise

LGM: All 4 Labs, all 3 Groups, SNR = 0dB

Linear Gaussian Scale

-0.42 0 0.42 0.84 1.26 1.68

Ident  SC1  SC2  BSS  MWF  COH

better
slightly better
very slightly better
very slightly better
slightly better
Preference Rating

Office room – 90/180/270° babble noise

LGM: All 4 Labs, all 3 Grops, SNR = 5dB

Linear Gaussian Scale

very slightly better
slightly better
better

Ident  SC1  SC2  BSS  MWF  COH
Preference Rating

Office room – 90/180/270° babble noise

LGM: All 4 Labs, all 3 Groups, SNR = 10dB

Linear Gaussian Scale

better
slightly better
very slightly better
very slightly better
slightly better
Comparison
LES / PrefRating / Intelligibility

Test Retest

• Preference rating: No significant difference (t-test, p>0.5)

• SRT: Significant difference (p<0.001), average retest scores are 0.2 dB higher as test scores

• Listening effort: Significant difference (p<0.001), average retest scores are 0.12 scales lower as test scores
Comparison
LES / PrefRating / Intelligibility

SNR range

Preference Rating:

Listening Effort: extrem middle very low

Intelligibility: 0 % 50 % 100 %

SNR: -10 dB 0 dB 10 dB 20 dB
Comparison
LES / PrefRating / Intelligibility

Speech Reception:
- very precise
- established
- most important feature
- needs time

Listening Effort:
- broader SNR range than speech intelligibility
- important feature, that helps to estimate cognitive demands

Preference Rating:
- even broader SNR range
- relative comparison
- works for very small differences
- overall feature (includes e.g. intelligibility, effort, artifacts etc)
The End