Subjective loudness ratings of vehicle noise with the hearing aid fitting methods NAL-NL2 and trueLOUDNESS

DIRK OETTING[1,3,*], JÖRG-HENDRIK BACH[1,2,3], MELANIE KRUEGER[1,2,3], MATTHIAS VORMANN[2,3], MICHAEL SCHULTE[2,3], AND MARKUS MEIS[1,2,3]

[1] *HörTech gGmbH, D-26129 Oldenburg, Germany*

[2] *Hörzentrum Oldenburg GmbH, D-26129 Oldenburg, Germany*

[3] *Cluster of Excellence Hearing4all, D-26129 Oldenburg, Germany*

Subjects with similar hearing thresholds showed large differences in loudness summation of binaural broadband signals after narrowband loudness compensation. Based on these findings, the fitting method trueLOUDNESS was developed to restore the individual binaural broadband loudness perception. In the present study, the trueLOUDNESS fitting method was compared with NAL-NL2. Loudness judgements of different vehicles' sounds were compared with average judgements by normal-hearing subjects. The loudness judgements with trueLOUDNESS fittings were closer to normal compared to the loudness judgements with NAL-NL2 fittings. This study shows that the lab measurement of binaural broadband loudness perception has validity beyond the laboratory.

INTRODUCTION

Individual loudness perception plays an important role in the fitting of hearing aids. The EuroTrak survey in Germany (EuroTrak DE, 2018) showed that the dimension "comfort with loud sounds" was the most important criterion for the overall satisfaction with hearing aids. In subjects with similar hearing thresholds the loudness summation of binaural broadband signals can differ substantially from the average. The effect has been well described (Oetting et al. 2016, 2018), and the amount of available data for the individual binaural broadband loudness summation of hearing-impaired (HI) subjects continue to increase. Based on the findings concerning the individual differences of binaural broadband loudness perception, the hearing-aid fitting method, trueLOUDNESS, was developed. The fitting rationale of trueLOUDNESS is to restore the binaural broadband loudness perception. The categorical loudness scaling measurements required for the trueLOUDNESS fitting are performed over headphones in the lab. The aim of this study was to show that the lab measurements of loudness scaling are related to real-world loudness perception with hearing aids. Therefore, a user group ("power users") for whom the trueLOUDNESS procedure predicts higher gains than NAL-NL2 was selected. Consequently, we expected that the loudness perception with an NAL-NL2 fitting would be lower than normal in this group. For a "sensitive user" group for whom the trueLOUDNESS procedure predicts lower gains than NAL-NL2, we expected that perceived loudness would be higher than normal with NAL-NL2 because NAL-NL2

*Corresponding author: d.oetting@hoertech.de

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 7: Auditory Learning in Biological and Artificial Systems, August 2019, Nyborg, Denmark. Edited by A. Kressner, J. Regev, J. C.-Dalsgaard, L. Tranebjærg, S. Santurette, and T. Dau. The Danavox Jubilee Foundation, 2019. ISSN: 2596-5522.

gains are above the gains required for restoring the loudness perception according to trueLOUDNESS.

METHOD

Screening

The subjects were selected from a pool of N = 129 subjects for whom the trueLOUDNESS gains from screening measurements were available. The subjects were older adults with an average age of 74 years (standard deviation: 6.5 years) with an averaged pure tone average (PTA) at 500, 1000, 2000 and 4000 Hz of 46.3 dB HL (range: 23.8 to 80.0 dB HL). Slightly more men (56%) than women were in the data set. Sixty-three procent of the subjects were hearing aid wearers. The average trueLOUDNESS gain at 500, 1000, 2000 and 4000 Hz was divided by the PTA for each ear. The values for the left and the right ear were averaged, resulting in a single-number value (average relative gain) for each subject. Fig. 1a shows for all 129 subjects the relative trueLOUDNESS gains over the PTA. The data show a large individual variation in terms of relative gains necessary to restore the loudness for a binaural broadband speech signal at 65 dB SPL even for subjects with similar PTAs.

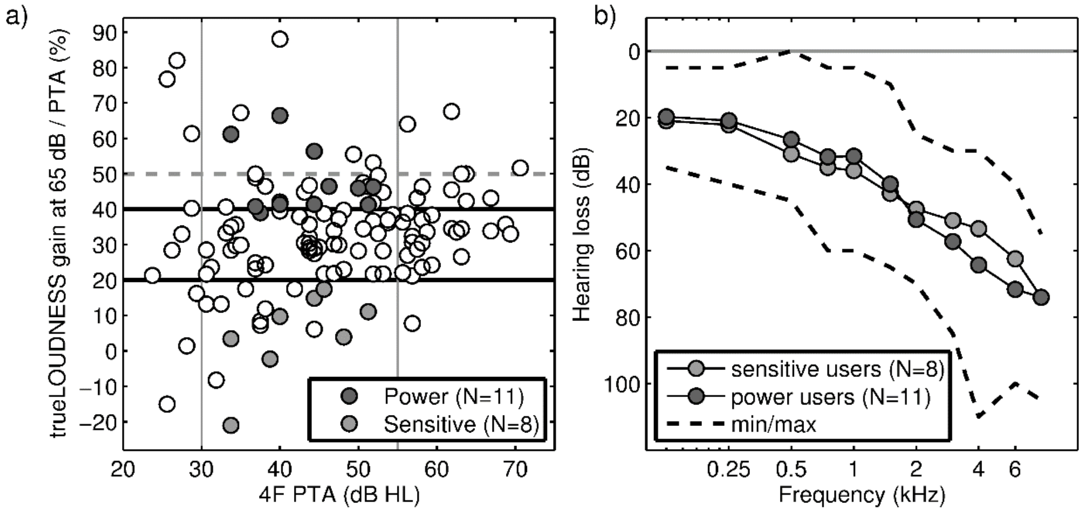

Fig. 1: a) Relative gain according to trueLOUDNESS for a binaural speech signal of 65 dB SPL for N=129 subjects. The average trueLOUDNESS gains at the PTA frequencies were divided by the averaged PTA for both ears. The grey dashed line indicates the half-gain rule at 50% relative gain. Subjects with relative gains below 20% are referred to as “sensitive users”, whereas “power users” have more than 40% relative gain. b) Averaged audiograms of the sensitive and power users that participated in this study. The dashed lines indicate the minimum and maximum values of all HI subjects.

For example, subjects with a PTA of around 40 dB HL showed relative gains between 0% and 90%. A value of 0% relative gain leads to an average of 0 dB gain at the four frequencies to restore binaural broadband loudness perception, and a value of 90%

relative gain leads to an average gain of 36 dB at the four frequencies. The difference of the relative gains between the left and the right ear were small, with values below 13.2 percentage points for 90% of the subjects. Five subjects showed differences of the relative gains for the left and the right ear above 20 percentage points, mostly due to asymmetric hearing loss.

Subjects

The subjects that participated in this study were selected to match the following inclusion criteria: 1) PTA averaged across both ears between 30 and 55 dB HL; 2) relative trueLOUDNESS gain at 65 dB SPL averaged across both ears below 20% or above 40%, respectively; 3) absolute difference of the PTA between the left and the right ear below 15 dB.

Subjects were selected with the expectation that for them gain prescription would differ between trueLOUDNESS and NAL-NL2. Relative gains for the standardized audiograms by Bisgaard et al. (2011) with PTAs between 30 and 55 dB HL (first inclusion criteria) were analysed. The relative gains of these audiograms were between 24% and 29%. Therefore, subjects with relative gains between 20% and 40% were excluded from this study. Subjects with relative trueLOUDNESS gains below 20% were assigned to a group referred to as "sensitive users" and subjects with relative trueLOUDNESS gains above 40% were assigned to a group referred to as "power users".

Eight sensitive users and 10 power users participated in this experiment. Details of both groups are shown in Table 1.

	Gender (f: female, m: male)	Age in years: mean (std.)	PTA in dB HL: mean (std.)	Experience with hearing aids
Sensitive users (N = 8)	5 f, 3 m	75.1 (3.5)	42.0 (6.5)	4 exp, 4 new
Power users (N = 10)	5 f, 5 m	74.7 (2.9)	42.6 (6.1)	6 exp, 4 new
Normal-hearing subjects	9 f, 10 m	50.7 (19.2)	5.2 (4.6)	–

Table 1: Comparison of the details of the subjects in the sensitive user group (below 20% relative gain, cf. Fig. 1) and the power user group (above 40% relative gain, cf. Fig. 1). The mean and standard deviation in age and PTA as well as the gender and experience with hearing aids were similar in both groups.

The maximum difference between the left and right relative trueLOUDNESS gains was 12.1 percentage points for the 18 subjects, indicating symmetric hearing losses. The PTA of the better ear was between 31.3 and 51.3 dB HL, which corresponds to the WHO criterion for hearing loss of grade 1 (slight impairment) and 2 (moderate impairment), respectively. The average hearing loss for both groups is shown in Fig. 1b. A total of 19 normal-hearing (NH) subjects (9 female, 10 male) participated as a reference group in this experiment. Their better-ear PTA was below 25 dB, which corresponds to the WHO criterion for grade 0 (no impairment). All participants were recruited through the database of Hörzentrum Oldenburg.

Location, vehicles and driving actions

The experiment took place at a former military airport in Oldenburg, where a set of roads was chosen to conduct the experiments with real vehicles and minimum disturbance. Four different vehicles were used: a car (Opel Corsa, 2016), a motorbike (Suzuki VX 800 800cc, 1994), a van (Fort Transit FT100, 1999) and a street sweeper (Kärcher MC 50).

The driving instructions for the first three vehicles were: idling (standing still with the engine on), accelerating, passing by at 30 km/h, passing by at 50 km/h, and breaking until standing still. The street sweeper's actions were standing by, standing by with the brushes switched on, and moving forward with the brushes switched on. Each driving action was repeated twice, once with the vehicle driving on the nearby side of the street with ~3 meters distance to the subjects and once with the vehicle driving on the far side of the street with ~6 meters distance to the subjects. Ten actions per vehicle and 6 actions for the street sweeper were conducted, resulting in 36 test conditions. Details on the driving actions including the recorded sound pressure levels are reported in Llorach et al. (2019). The subjects were placed parallel to the road. Seven positions were available on the left side of a central recording station and seven places on the right side.

Fitting of hearing aids

The HI subjects were equipped with Phonak Audéo B90-312 hearing aids. For the acoustic coupling individual ear moulds (cShells, if available) or domes were used. The type of dome (open, closed, or power dome) was selected according to the recommendation by the Phonak fitting software. Two different programs in the hearing aids were set up. In program 1 the gains were adjusted according to the trueLOUDNESS gain calculations for 50, 65, and 80 dB SPL. The trueLOUDNESS target gain functions were visible in the fitting software. The gains were adjusted manually by an acoustician to achieve a close match between the target trueLOUDNESS functions and the gain functions of the hearing aid. In program 2 the fitting method NAL-NL2 was selected as the gain calculation method in the fitting software. The gender and experience-specific differences were considered during fitting, and the gain level was adjusted to 100% target gain.

Procedure

The subjects reported for each condition their loudness perception on a printed scale of the categorical loudness scaling procedure from "not heard" to "too loud" (Brand and Hohmann, 2002). The vehicle's starting point was in front of the recording devices. The experimenter showed numbers to indicate the current action. The subjects were asked to rate the loudness and annoyance of that action. Each run of a vehicle took between 7 to 8 minutes to complete all driving actions. The vehicle order was car, motorbike, van, street sweeper. First, program 1 (trueLOUDNESS) was tested with all cars and all driving actions; then program 2 (NAL-NL2) was tested. A break was included between the programs. NH subjects participated and performed the loudness and annoyance ratings at the same time as the HI subjects. Each session lasted in total around 1.5 h. Overall, four sessions on two different days were conducted. On the first day, sessions 1–3 were conducted, and on the second day, session 4 was conducted. The participants per session are shown in Table 2. Calibrated sound recordings have been made to ensure similar test conditions across all sessions, retrospectively. An ethics proposal was approved by the ethical commission of the University of Oldenburg.

Session:	1	2	3	4	Total
HI “sensitive users”	5	1	0	2	8
HI “power users”	1	1	4	4	10
NH	0	10	5	4	19

Table 2: Participants of the different groups over the different sessions of the experiment. Four sessions were conducted. Session 1, 2 and 3 took place on the same day. Session 4 was conducted about one month later.

RESULTS

To compare the NH and HI results, the median loudness responses from the categorical loudness scale in categorical units (CU) were calculated. For the NH subjects, the sensitive users and the power users, the median value for each of the 36 conditions was calculated. Each condition was assigned to the NH loudness category between 0 and 50 CU in 5 CU steps that was nearest to the median NH value. For each category, the median of the assigned condition was calculated for the sensitive and power users with the NAL-NL2 and the trueLOUDNESS fitting. The results are shown in Fig. 2. The abscissa shows the NH loudness category along with the number of conditions assigned to that category. E.g., there were 13 conditions that resulted in a median NH loudness rating of 25 CU (“medium loud”) and 2 conditions that resulted in a median NH loudness rating of 45 CU (“very loud”). If the resulting loudness ratings with hearing aids corresponded to the NH loudness ratings within an error margin of ±2.5 CU, the values would be inside the white diagonal corridor. The resulting median values for the sensitive users with NAL-NL2 fitting are above the diagonal line. For the power users with NAL-NL2 fitting, the median loudness values are below the diagonal line for loudness categories of 30 CU and above. The median values for the sensitive and the power groups with the trueLOUDNESS fitting are close to the diagonal line. That means that for the power users, the trueLOUDNESS fitting predicts higher gains than NAL-NL2, which shifts the resulting loudness ratings towards the diagonal line. For the sensitive users, the trueLOUDNESS fitting predicts less gain than NAL-NL2, which again shifts the resulting loudness ratings closer to the diagonal line.

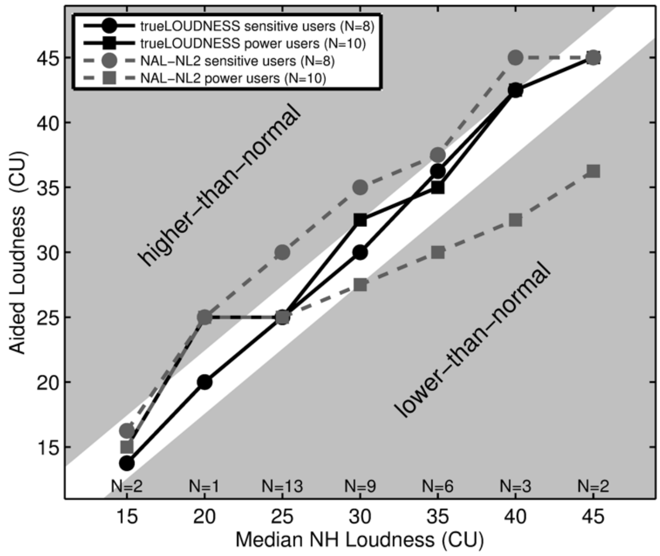

Fig. 2: The test conditions were pooled according to the median NH loudness rating and assigned to the category closest on the loudness scale from 0 to 50 CU in 5 CU steps. The four lines show the median results of the sensitive and power users after fitting with trueLOUDNESS and with NAL-NL2.

For the statistical analysis of the data, the loudness ratings were transformed to the phon scale (interval scale) using the transformation by Heeren et al. (2013) and an interpolation of the tabulated values according to ANSI S3.4 (2007). The difference between the power and sensitive users, and the average NH value was calculated and is referred to as *mean Δloudness re. NH*. These differences for each condition indicate the deviation from the mean NH rating. Fig. 3a shows the *mean Δloudness re. NH* over the *mean NH loudness value in phons* for the trueLOUDNESS fitting.

The conditions with values around 40 phons were the idle conditions of the car (near and far side) rated by the NH subjects on average with 37.6 and 40.1 phons. The regression lines show on average a slightly increased loudness rating in the power users for the trueLOUDNESS fitting. For NH mean values above 75 phons, the deviation according to the regression line was less than 3 phons in both HI groups. The results for the NAL-NL2 fitting are shown in Fig. 3b. The regression line for the sensitive users (grey crosses) show constantly higher loudness ratings for all conditions from low to high NH mean values. The power users with the NAL-NL2 fitting showed higher loudness ratings from 2 phons for the conditions with low NH mean values and −8.8 phons for the conditions with high NH mean values. The *mean Δloudness re. NH* for each group and each fitting were pooled and are shown by the boxplots in Fig. 3c. Whiskers mark values within 1.5 interquartile ranges of the first and third quartile, respectively. Outliers are marked with a cross.

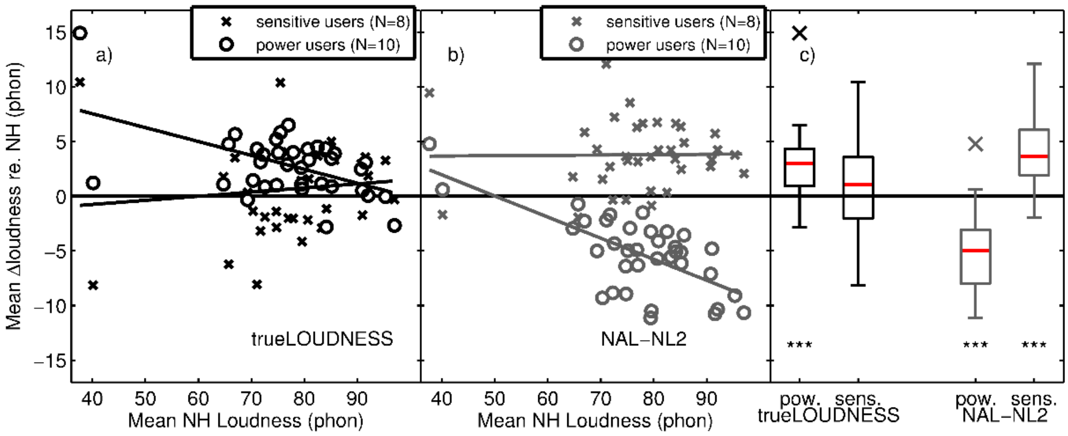

Fig. 3: Distance from the NH mean loudness rating in phons over the mean NH loudness rating in phons for both groups when fitting according to a) trueLOUDNESS and b) NAL-NL2. c) Pooled data over all test conditions. Median values for trueLOUDNESS were closer to zero compared to NAL-NL2 for both groups.

The median values for trueLOUDNESS were 3.0 phons and 1.1 phons for the power users and for the sensitive users, respectively. The median values for NAL-NL2 were −5.0 phons and 3.6 phons for the power users and for the sensitive users, respectively. For each of the four pooled data sets, we performed a Wilcoxon signed rank test with the null hypothesis that the data come from a distribution whose median is zero against the alternative hypothesis that the distribution does not have zero median. Only for the sensitive users with trueLOUDNESS fitting, the test indicated that the null hypothesis cannot be rejected ($p = 0.61$). For the three other conditions, the test indicated a median value different from zero ($p < 0.001$).

DISCUSSION AND CONCLUSION

In this experiment we tested whether the trueLOUDNESS measurements from the lab are related to the real-world loudness perception with hearing aids. We selected a user group for whom the trueLOUDNESS procedure predicts higher gains than NAL-NL2 ("power users"). The results from Fig. 3c showed that in this group the deviation from normal loudness with NAL-NL2 was between −3.1 and −8.0 phons (interquartile range), indicating that loudness was lower than normal in the power user group.

For the sensitive user group, we expected that perceived loudness would be higher than normal with NAL-NL2. The results in Fig. 3c showed a higher-than-normal loudness perception of 2.0 to 6.1 phons (interquartile range) with NAL-NL2. Both groups clearly separate in Fig. 3c, indicating that the grouping according to the trueLOUDNESS measurements in the lab reflects real-world loudness perception with hearing aids. Using the trueLOUDNESS gain prescription, normal loudness perception was achieved in the sensitive user group on average. For the power users, trueLOUDNESS gain predictions lead to higher-than-normal ratings of about 0.9 to 4.3 phons (interquartile range). These results can be used to define a gain reduction rule for reducing the amount of loudness compensation for power users with the trueLOUDNESS fitting procedure.

The following conclusions can be drawn:

- Binaural broadband loudness scaling results from the lab according to trueLOUDNESS are related to the real-world loudness perception with hearing aids. This indicates that the binaural broadband loudness scaling results are reliable indicators during fitting of hearing aids to avoid under- or over-amplification.
- Hearing aids fitted with NAL-NL2 result in higher-than-normal loudness ratings for real-world vehicle sounds for one group of users (sensitive users), and lower-than-normal ratings for a second group of users (power users). These groups were defined based on their trueLOUDNESS-based relative gains, which shows that trueLOUDNESS measurements from the lab can be used to identify these groups of subjects.
- Prescription rules based solely on the hearing threshold cannot be further tuned towards a better loudness match for sensitive and power users simultaneously. Further information from binaural broadband loudness perception is required to decide if gains should be reduced or increased for an individual listener.
- The trueLOUDNESS fitting leads on average to a normal loudness perception in sensitive users and slightly higher-than-normal perception in power users.

REFERENCES

ANSI S3.4 (**2007**). American national standard: Procedure for the computation of loudness of steady sounds. American National Standards Institute.

Brand, T. and Hohmann, V. (**2002**). "An adaptive procedure for categorical loudness scaling," J. Acoust. Soc. Am., 112, 1597–1604, doi: 10.1121/1.1502902

EuroTrak DE (**2018**). EuroTrak DE 2018. Retrieved from https://www.ehima.com/wp-content/uploads/2018/06/EuroTrak_2018_GERMANY.pdf on 30.08.2019.

Heeren, W., Hohmann, V., Appell, J.-E., and Verhey, J.L. (**2013**). "Relation between loudness in categorical units and loudness in phons and sones," J. Acoust. Soc. Am., 133, 314–319, doi: 10.1121/1.4795217

Llorach, G., Oetting, D., Krueger, M., et al. (**2019**). "Vehicle noise: Loudness ratings, loudness models and future experiments with audiovisual immersive simulations," Proc. Inter-Noise 2019, Madrid, Spain.

Oetting, D., Hohmann, V., Appell, J.-E., Kollmeier, B., and Ewert, S.D. (**2016**). "Spectral and binaural loudness summation for hearing-impaired listeners," Hear. Res., 335, 179–192, doi: 10.1016/j.heares.2016.03.010

Oetting, D., Hohmann, V., Appell, J.-E., Kollmeier, B., and Ewert, S.D. (**2018**). "Restoring perceived loudness for listeners with hearing loss," Ear Hear. 39, 664–678, doi: 10.1097/AUD.0000000000000521