MADRID
inter.noise 2019
June 16 - 19
NOISE CONTROL FOR A BETTER ENVIRONMENT

Vehicle Noise: Loudness Ratings, Loudness Models and Future Experiments with Audiovisual Immersive Simulations

Llorach, Gerard[1,3,4] **Vormann, Matthias**[1,3] **Hohmann, Volker**[1,2,3]
Oetting, Dirk[2,3] **Fitschen, Christina**[2,3] **Meis, Markus**[1,2,3]
Krüger, Melanie[1,2,3] **Schulte, Michael**[1,3]

[1]Hörzentrum Oldenburg, [2]HörTech, [3]Cluster of Excellence Hearing4all, [4]Medizinische Physik, Universität Oldenburg
Marie-Curie-Straße 2
26129 Oldenburg, Germany

ABSTRACT

Loudness complaints are still very common among hearing-aid users. Therefore, loudness is an important issue that needs to be addressed in hearing aid research to improve and optimize hearing aid fitting. Empirical data show that only 2/3 of the hearing-aid users are satisfied with the comfort of loud sounds. The topic itself is quite challenging, as loudness is a subjective measure and it can be perceived differently depending on the context and conditions. To extend the research between laboratory and field presentation, we collected the ratings of loudness and annoyance of four different land vehicles (car, van, motorcycle and street sweeper) with different speeds and behaviours in normal-hearing and hearing-impaired listeners. In this work we present and analyse the ratings of 15 normal-hearing participants and compare their average field ratings to the calculated loudness with loudness models. Furthermore, possible future immersive audiovisual experiments for lab vs reality loudness comparisons are presented.

Keywords: Loudness, Annoyance, Vehicles, hearing aid
I-INCE Classification of Subject Number: 79
http://i-ince.org/files/data/classification.pdf

1. INTRODUCTION

Loudness and annoyance have been studied for many years, resulting into a series of standards and normalizations [1, 2]. Most of the experiments and standards are only considering acoustic stimuli in controlled laboratory conditions, which is far from a realistic situation. Nevertheless, research showed that adding visual information and doing more realistic laboratory simulations can change the perceived loudness and annoyance of sounds. A known example is the experiment [3] with the red and green train, where red trains were perceived louder than green trains, although the acoustic stimuli were the same. Loudness ratings were even more reduced when being in a more realistic experience such as a car simulator [3].

Corresponding author email: gerard.llorach.to@uni-oldenburg.de

Still, standards have been developed in audio-only laboratories. And even less research has been done comparing loudness and annoyance in the field and in laboratory settings. In [4], normal-hearing (NH) and hearing-impaired (HI) participants wore hearing aids in their normal live for a week with the possibility to adjust the volume and record loudness levels. NH listeners preferred realistic loudness levels in the field whereas hearing-impaired participants adjusted their devices to lower-than-normal loudness. The same subjects participated in an audiovisual laboratory experiment where different real-life listening situations with different loudness levels were presented and they had to rate the loudness with their hearing aids. Both groups now preferred lower-than-normal loudness settings i.e. NH listeners perceived higher levels of loudness in the laboratory.

One of the reasons for the results of [4] in the laboratory could be that the presentation of the stimuli was on a display screen with stereo audio. As mentioned in [3], loudness perception could be reduced when adding more realism and immersion to the simulation. Thus, participants might choose normal loudness settings instead of reducing it in the laboratory when immersive realistic audiovisual simulations are used, proving to be a more ecologically valid method. This is one of the focuses of the study presented here: to provide a comparison of the ratings in the field and in the laboratory. The second focus is to examine the relationship between subjective ratings and objective measures of loudness and annoyance, which have only been established in laboratory settings [2, 5, 6].

In our work, ratings from the field were collected to be later compared to different laboratory conditions. Four different land vehicles were driven in an empty street with different driving actions such as stopping at a red light or passing by at 30 or 50 km/h. Participants rated loudness and annoyance of the actions of the vehicles. These driving actions were all recorded with a 360º camera and a tetrahedral microphone, in order to be able to produce different conditions in the laboratory. In this paper we present the field experiment procedure, the NH field data compared to loudness models, and introduce the experiments planned in the laboratory.

2. METHOD

The goal of the study was to collect ratings in the field to later compare them to the exact same stimuli/situation in the laboratory. Doing measurements in the field can be problematic, as the situations are less controllable, and many variables can influence the ratings of loudness and annoyance. A street on a former military area in Oldenburg was used for the experiments as no traffic is permitted without previous admission. Each one at a time, four different vehicles (car, motorcycle, van and street sweeper) were driven with controlled driving actions. Participants were sitting on the side of the road and rated each driving action with a categorical loudness scale (CLS) [7] and an ICBEN numerical annoyance scale (0-10) [8].

The experiment in the field was recorded with a 360º camera (Xiaomi Mi Sphere Camera), a tetrahedral microphone (Core Sound TetraMic) and a level meter, placed in the middle of the row of participants (Figure 1). With these recordings, one can reproduce immersive audiovisual stimuli: the 360º video recordings can be reproduced through head-mounted displays or other immersive displays; and the tetrahedral microphone recordings can be used for first-order Ambisonic (FOA) reproduction or to render other audio formats (stereo, mono).

2.1 Participants

Corresponding author email: gerard.llorach.to@uni-oldenburg.de

19 NH and 20 HI listeners participated in the study (Table 1). The data was collected in four sub-groups to handle sitting places, time for hearing-aid fitting and to conduct and guide the experiment. The participants were recruited through the database of Hörzentrum Oldenburg.

Table 1. Gender, age and PTA of the NH participants of the experiment.

	Female / Male	**Mean age (SD)**	**PTA (SD)**
19 normal-hearing	9 female, 10 male	50 (19.2) years	3.8 (4.7) dB HL

2.2 Vehicles and driving actions

Four different vehicles were used in this experiment: a white car (Opel Corsa 2016), a red motorbike (Suzuki VX 800 800cc 1994), a dark blue van (Fort Transit FT100 1999) and a street sweeper (Kärcher MC 50). The driving instructions for the first three vehicles were: standing by (stopped with the engine on), accelerating, passing by at 30 km/h, passing by at 50 km/h and breaking until stopping. The street sweeper's actions were standing by, standing by with the brushes on and moving forward with the brushes on. Each driving action was repeated twice, once from a close distance (3 meters) and once from a far distance (6 meters) and from different directions (left-right or right-left) (Figure 1), resulting into 10 actions per vehicle and 6 actions for the street sweeper. The driving actions are shown in Table 2 together with their maximum dB SPLs calculated with a 125ms window. These driving actions were chosen in resemblance of the typical ones in an urban area.

Table 2. Vehicle's driving actions with average max dB SPL. The actions are numbered with the order of presentation during the experiment. LR and RL stands for the direction of the driving: Left-to-Right (LR) and Right-to-Left (RL).

	Mean max dB SPL of the driving actions									
	1A. Stand by (close)	**2A. Accelerate LR (close)**	**3A. 30 km/h RL (far)**	**4A. 50 km/h LR (close)**	**5A. Break and stop RL (far)**	**6A. Stand by (far)**	**7A. Accelerate RL (far)**	**8A. 30 km/h LR (close)**	**9A. 50 km/h RL (far)**	**10A. Break and stop LR (close)**
Car	71.2	84.3	73.3	81.5	75.2	67.9	80.1	75.2	76.9	77.1
Motorbike	83.5	91.5	82.5	89.7	81.1	78.4	86.6	89.0	88.1	84.0
Van	82.7	88.4	81.1	90.1	80.5	80.3	87.8	84.5	85.9	82.8
	1B. Stand by (close)	**2B. Brushes on (close)**	**3B. Forward LR (close)**	**4B. Stand by (far)**	**5B. Brushes on (far)**	**6B. Forward RL (far)**				
Street sweeper	83.6	91.1	92.6	76.9	83.7	83.5				

Each vehicle did the driving actions in two different sessions. For the NH listeners the sessions were a test and a retest session. The HI listeners used a different hearing-aid fitting method for each session. The vehicles were driven in the following order: car, motorbike, van and street sweeper (test); street sweeper, car, motorbike and van (retest). The order of the driving actions is presented in Table 2.

Corresponding author email: gerard.llorach.to@uni-oldenburg.de

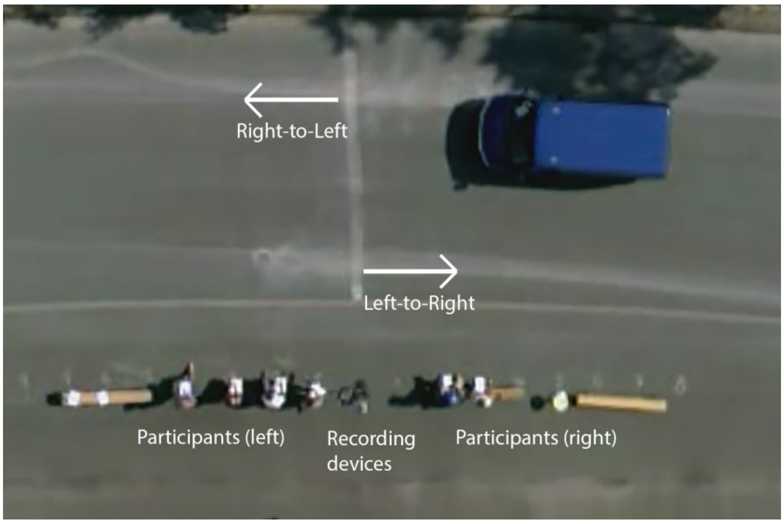

Figure 1. Top view of the street and sitting positions of the participants during the experiment.

2.3 Procedure

Participants filled in the sound preference and hearing habits questionnaire [13]. They were indicated to sit on the benches or chairs placed on the side of the road. The loudness and annoyance questionnaires were handed out and instructions given for the experiment. The vehicle's starting point was in front of the recording devices and the participants were sited on the same side of the road on both sides of the recording position (Figure 1). The experimenter showed numbers to indicate which action was happening and the participants were asked to rate the loudness and annoyance of that action. Each vehicle took between 7 to 8 minutes to complete all driving actions. Between the test and retest participants could take a rest. Each session lasted in total around 1.5h.

2.4 Loudness models

The ANSI S3.4 (2007) [9] was used to calculate the loudness (sone) for the stationary sounds (in Table 2 actions 1A, 6A, 1-2B, 4-5B) such as standing by. The N5 percentile loudness measure from [2] was used for the loudness (sone) of the time-varying sounds (in Table 2 actions 2-5A, 7-10A, 3B, 6B) such as accelerating and passing by. The LoudnessToolbox 1.2 for Matlab provided by [10] was used to calculate the loudness levels.

In our experiment, each participant was sitting in a different position and was at a different distance and view perspective of the vehicle (Figure 1). Thus the sound levels and visual stimuli were not the same for the participants and the recording devices (camera and microphones). We compensated for the individual sound levels and the individual calculated loudness according to the inverse proportional law using the position of the microphone, the position and driving direction of the sound source and the position of each participant. For example, participants sitting at the outer positions (~6 meters from source) received ~6 dB SPL less than the recording microphone (~3 meters from source) when the vehicle was standing by. When the vehicle was accelerating from the middle position (in Table 2 actions 2A, 7A, 3B, 6B) or breaking and stopping (in Table 2 actions 5A, 10A), we only compensated for the participants that would not see the vehicle passing in front of them i.e. if the vehicle was accelerating from left to right, the participants on the left would experience lower levels than those recorded by the microphone. When the vehicles were passing by at 30km/h

---

Corresponding author email: gerard.llorach.to@uni-oldenburg.de

or 50km/h (in Table 2 actions 3-4A, 8-9A) the sound levels and calculated loudness of the position of the microphone were used, as they were the same for all participants and recording devices.

3. RESULTS

In this section we show the subjective ratings of loudness and annoyance of the NH listeners and describe the relationship with the objective measures and models of loudness. We also show the variability of the driving actions and the test re-test reliability.

3.1 Loudness Ratings and Loudness Models

The loudness ratings in categorical units (CU) can be seen in Figure 2. The CUs are plotted against their equivalent dB SPL and sone levels. We fitted the parameters of equation (Eq. 1) to the medians of each loudness category to establish the relationship between CUs and sones according to previous work by [5, 6]:

Equation 1 $CU = -32.16 \log(sone)^3 - 151.15 \log(sone)^2 - 181.0 \log(sone) + 72.63$

The fitted curve (dashed red line) had a higher slope (Figure 2.b) compared to the literature results (blue and green line): at low loudness values the results of the field-studied showed lower loudness ratings in the field than expected by the transformation formula calculated in the laboratory [5, 6]. The root mean square error (RMSE) was 7.68 CU. This values was higher in comparison to 0.75 CU and 0.30 CU from [5, 6] respectively. In [5, 6] stationary narrow-band noises were used in headphone experiments whereas we had binaural broadband time-varying stimuli with visual information measured in realistic situations. In Figure 2a) the mean (green solid line) and the standard deviation (green dashed line) of the loudness ratings are shown.

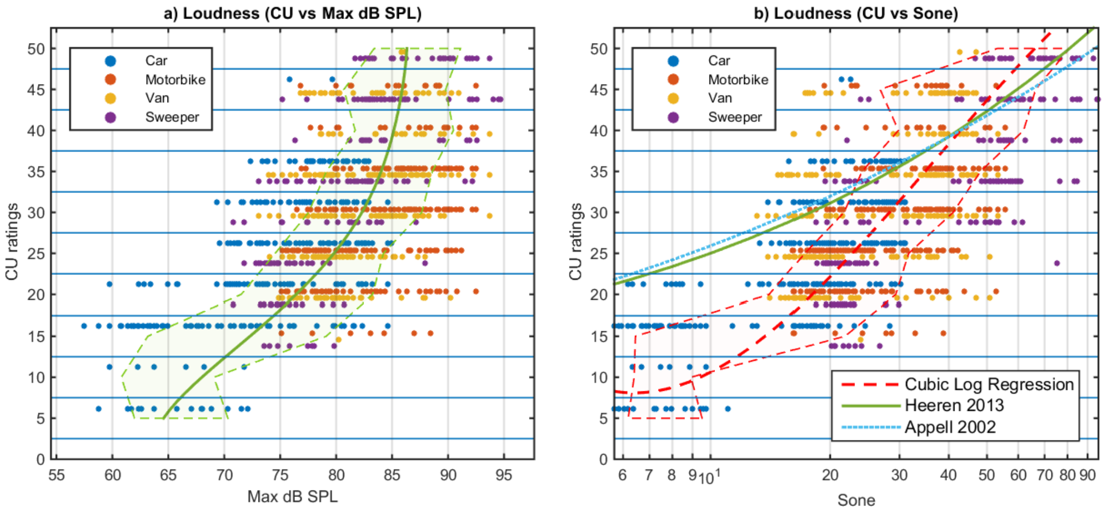

Figure 2. Ratings of the CLS in Categorical Units (CU) against the maximum dB SPL (a. - left) and the loudness levels (b. - right) of each driving action. The models of loudness proposed by Heeren et al. 2013 (green) [6] and Appell 2002 (blue dotted) [5] and the new fitted curved (red dashed) are also shown (b. - right).

3.2 Annoyance ratings

It was shown that annoyance correlates with loudness (sone) [11] for urban noises (airplane, road and train). In our experiment the relationship between the

---

Corresponding author email: gerard.llorach.to@uni-oldenburg.de

annoyance ratings (Pearson's correlation) was higher with the subjective loudness CU ratings (r = 0.820) and smaller for the computed loudness in sones (r = 0.578) and the maximum dB SPL levels (r = 0.472). However, all correlations were significant (p<0.001).

3.3 Driving variability and Individual variability

Each driving action (Table 2) was repeated 8 times (4 sub-groups, test and re-test). The vehicles were driven by an experimenter, thus the driving behavior was not identical across repetitions. The driving actions had an average standard deviation of 1.74 dB SPL (range 0.52-4.65 dB SPL) and an average standard deviation of 2.74 sone (range 0.55-12.84 sone). The test-retest variability is shown in Figure 3.

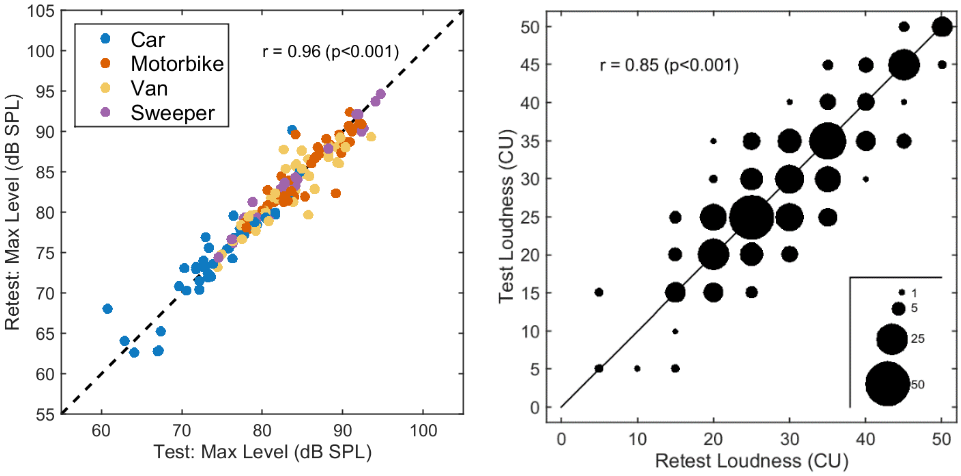

Figure 3. Test-retest reliability of the driving actions (left) and the loudness ratings (right).

4. DISCUSSION

Loudness ratings in the field were lower than expected by the loudness models. This could probably be due to the fact that the loudness models were mainly evaluated in the laboratory and not in the field. Thus, it could be hypothesized that loudness in audio-only laboratories is rated louder than it is in a realistic situation. This is also supported by the findings of [4], where the participants chose lower gains in the laboratory with a video screen in comparison to the field. These differences between the field and the laboratory could be for different factors. Visual cues play an important role, mostly reducing loudness perception [3]. The loudness models used in this experiment [5, 6] were created in audio-only conditions, thus probably inducing higher loudness ratings. Acoustic context could also be important: presenting stimuli in an anechoic environment is quite different from doing it in the field where background noise is usually present and sounds can be separated by incident directions. The scene in the field could give some context and loudness reference to the participants that might not be available in the laboratory experiments.

In our field experiment some of the requirements for precise loudness ratings could not be fulfilled [7]. For example, stimuli could not be repeated systematically but only with similar levels (far and close) i.e. for each stimuli (driving action) we only had four ratings per subject at two different levels (test-retest, far-close). Additionally we didn't cover all the dynamic range of the participant (from "not heard" to "extremely

---

Corresponding author email: gerard.llorach.to@uni-oldenburg.de

loud"). Additional measurements with other land vehicles, such as a bicycle, could be done in future experiments to cover all the dynamic range.

Annoyance ratings were found to be correlated with the loudness ratings. The correlation coefficient was not as high as expected [11] for the objective measures of loudness. Additional objective measures such as psychoacoustic annoyance, sharpness, roughness and fluctuation could be included in future work.

5. POSSIBLE LABORATORY EXPERIMENTS

Before the field experiments, microphones and a GPS tracker were attached to the vehicles in order to record the sounds of the wheels and the engine relative to their position and velocity. Furthermore, the 360º video recordings and the FOA microphone recording of the field experiments are available. The objective of these recordings is to test the participants in different laboratory conditions and observe how the perception could be affected.

The conditions in the laboratory could vary, from the simpler to the most complex in order to see the effects of the loudness ratings and what is needed to achieve similar ratings from those in the field. Acoustic stimuli could be reproduced with mono, stereo, first-order Ambisonics or even up-sampled Ambisonics. Visual cues could be the video recordings, played over immersive displays (head-mounted display, surrounding screens), display screens or even left out (audio-only). Another approach would be to reconstruct the field scenes synthetically with 3D modelling, game engines and virtual acoustic rendering using the recordings of the attached microphones. With this virtual reconstruction one would gain more specific control of the stimuli (i.e. changing the color and shape of a vehicle, modifying the speed...). For a more detailed description of different methods of capture, synthesis and reproduction in the laboratory please refer to [12].

5.1 Example 1. Reality Replication

The goal of this experiment would be to achieve the necessary realism of a simulation to get the same ratings as in the field. The same stimuli recorded in the field would be reproduced in the laboratory, with the same acoustic levels. The methods chosen would try to achieve the maximum fidelity to the field. In our case that would be immersive displays (360º videos) with surrounding audio (first-order Ambisonic recordings). For each laboratory condition the transformation between levels and loudness ratings could be calculated. These transformations would be then compared with the one obtained in the field (Eq. 1) in order to find out which level of realism matches best the field ratings.

The stimuli could be presented with two different approaches. One would be a reality-vs-laboratory approach, where the subject would be throughout the whole experiment in the virtual environment in order to avoid context effects (background noise and background scene constantly running between stimuli). The other approach would be a clinical one: only the stimuli to be rated would be presented and silence would be present between stimuli.

5.2 Example 2. Level Adjustment

In this experiment the participants would have to adjust the acoustic levels themselves, in order to find which level would fit the stimuli presented. The goal would be to check how different laboratory conditions change these level adjustments and if they are close to the field levels.

Corresponding author email: gerard.llorach.to@uni-oldenburg.de

Many other experiments could be designed by distorting and modifying the stimuli recorded in the field i.e. changing the color of the vehicles, modifying the acoustic signals to be more pleasant, adding background noise, etc.

6. CONCLUSIONS

In this work we presented the loudness and annoyance ratings of a controlled real-life situation. Loudness ratings in the field were found to be lower than expected by loudness models. Annoyance ratings were found to be correlated with the loudness ratings i.e. the higher the loudness rating, the higher the annoyance rating. Most of the driving actions were repeated with little variation of the acoustic levels and the participants rated consistently the driving actions. Several possible laboratory experiments were introduced to compare the field ratings against different laboratory conditions.

7. ACKNOWLEDGEMENTS

This work has received funding from the EU's H2020 research and innovation programme under the MSCA GA 675324 (ENRICH) and the Deutsche Forschungsgemeinschaft (DFG, Cluster of Excellence EXC 1077/1 "Hearing4all" and SFB1330 Project B1 and C4).

8. REFERENCES

1. Scheuren, J., 2014, October. "*ISO 532-Living and working with alternative loudness standards"*. In INTER-NOISE and NOISE-CON Congress and Conference Proceedings (Vol. 249, No. 1, pp. 6029-6034). Institute of Noise Control Engineering.

2. Zwicker, E. and Fastl, H., 2013. "*Psychoacoustics: Facts and models"* (Vol. 22). Springer Science & Business Media.

3. Fastl, H., 2004. "*Audio-visual interactions in loudness evaluation*". In Proc. of Proc. Int. Congress on Acoustics ICA 2004, 18. Intern. Congress on Acoustics, Kyoto, Japan.

4. Smeds, K., Keidser, G., Zakis, J., Dillon, H., Leijon, A., Grant, F., Convery, E. and Brew, C., 2006. "*Preferred overall loudness. II: Listening through hearing aids in field and laboratory tests*". International Journal of Audiology, 45(1), pp.12-25.

5. Appell, J.E., 2002. "*Loudness models for rehabilitative audiology*" (Doctoral dissertation, Universität Oldenburg).

6. Heeren, W., Hohmann, V., Appell, J.E. and Verhey, J.L., 2013. "*Relation between loudness in categorical units and loudness in phons and sones*". The Journal of the Acoustical Society of America, 133(4), pp.EL314-EL319.

7. ISO, 2006. "*Acoustics Loudness scaling by means of categories*" *(ISO 16832:2006)*. International Organization for Standardization.

8. ISO, 2003. "*Acoustics Assessment of Noise Annoyance by Means of Social and Socio-acoustic Surveys" (ISO/TS 15666:2003)*. International Organization for Standardization.

9. ANSI, 2007. "*Procedure for the computation of loudness of steady sounds (ANSI S3. 4-2007)."* American National Standards Institute

10. GENESIS. *Loudness Toolbox 1.2.* <http://genesis-acoustics.com/sonie_en_ligne-32.html>. Last accessed: March 2019

11. Preis, A., Hafke-Dyx, H., Szychowska, M., Kocinski, J. and Felcyn, J., 2016. *Audio-visual interaction of environmental noise*. Noise Control Engineering Journal, 64(1), pp.34-43.

12. Llorach, G., Grimm, G., Hendrikse, M.M. and Hohmann, V., 2018, October. *Towards Realistic Immersive Audiovisual Simulations for Hearing Research: Capture, Virtual Scenes and Reproduction*. In Proceedings of the 2018 Workshop on Audio-Visual Scene Understanding for Immersive Multimedia (pp. 33-40). ACM.

13. Meis, M., Huber, R., Fischer, R.L., Schulte, M., Spilski, J. and Meister, H., 2018. "*Development and psychometric properties of the sound preference and hearing habits questionnaire (SP-HHQ)*". International journal of audiology, 57(sup3), pp.S118-S129.

Corresponding author email: gerard.llorach.to@uni-oldenburg.de