Overall loudness of short time-varying sounds

Josef Schlittenlacher[1]; Takeo Hashimoto[2]; Sonoko Kuwano[3]; Seiichiro Namba[4]

[1] Technische Universität Darmstadt, Germany

[2] Seikei University, Japan

[3,4] Osaka University, Japan

ABSTRACT

When predicting the overall loudness of time-varying sounds, it is necessary to use a function which calculates an overall value based on instantaneous loudness. Although N5 and LL(P) correlate for many sounds, their concepts are rather different. N5 is the loudness which is exceeded in five percent of the time, thus it is an ordinal value. LL(P) is the energetic mean of all instantaneous loudness values. For the present set of experiments, sounds with orthogonal conditions were chosen. This means they vary in N5 in a dynamic range of about 30 phon but have constant loudness according to LL(P) or vice versa. Participants judged the overall and instantaneous loudness of the sounds with ten seconds duration by magnitude estimation using line length and continuous judgment. The results show that LL(P) has some advantages compared to N5. In contrast to N5, it correlates highly with the overall loudness for the present set of stimuli.

Keywords: Loudness, Time-varying sounds   I-INCE Classification of Subjects Number(s): 63.1

1. INTRODUCTION

Although sounds in daily life usually vary significantly over time, it is quite easy for a listener to assign a single representative loudness rating to them. This single value is called overall loudness. Some of many examples for judging overall loudness in real life, outside the laboratory, are judging the loudness of a noisy environment or adjusting the loudness of the TV. There are several models which can predict such ratings of overall loudness, however, they use different assumptions how loudness is averaged over time. DIN 45631/A1 [1] proposes the N5, which is the loudness that is exceeded in five percent of time. Thus, it is an ordinal value which often is close to, but smaller than the maximum loudness. A different approach is the LL(P) [2]. Like the L_{eq} calculates the energy mean of sound pressure levels over time, the LL(P) calculates the energy mean of loudness levels over time. The concept of the LL(P) could be based on several loudness models for stationary loudness. In the following, it refers to the version which is based on the Zwicker model [3]. For the present paper, the terms LL(P) and LLz(P) are used equivalently..

Despite of the fact that N5 and LL(P) show a high correlation for many sounds in daily life, their concepts are completely different and that is why they sometimes predict remarkably different overall loudness. For stationary sounds, their predictions are the same. However, imagine an extreme example, in which the softer 95 percent are very soft compared to the loudest five percent. This does not change the N5, however, the LL(P) is decreased by up to 13 phon. On the other hand, imagine a rather stationary sound with an impulse or loud part shorter than five percent of the time. The N5 then reflects the stationary part, however, the LL(P) is higher because of the loud part.

It should be mentioned that there are further models for calculating the loudness of time-varying sounds. The LL(E) is similar to the LL(P), however, it first averages third-octave levels over time and afterwards calculates loudness. Because of their high correlation, only the LL(P) is considered in this study. The Glasberg and Moore model for the loudness of time-varying sounds [4] predicts several functions for loudness over time in order to represent instantaneous loudness on the one hand and memory effects on the other hand. However, it does not suggest how to take a single value for overall

[1] schlittenlacher@psychologie.tu-darmstadt.de
[2] gakuencho-hashimoto@board.seikei.ac.jp
[3] kuwano@see.eng.osaka-u.ac.jp
[4] qzw00041@nifty.com

loudness. For this reason, it will not be considered in this paper which focuses on overall loudness.

For the comparison of N5 and LL(P), it is desirable to find stimuli which result in orthogonal conditions between the two metrics. This is difficult for short stimuli because of the different implementations of time constants. While the model of LL(P) calculates the loudness level each 125 ms, DIN 45631/A1 uses time constants which are dependent on the level and duration of a preceding impulse or loud segment. However, an experiment with short stimuli has the advantage that more stimuli can be presented and that they can be repeated in several trials. A skeptic might point out that short durations are different from situations of typical noise immissions, however, the models do not make limitations to the duration, claiming to correctly predict the loudness of sound bursts with just some milliseconds duration ranging to sounds lasting several minutes or hours.

The present study consists of two experiments with the same set of stimuli. Real sounds with durations of ten seconds were found, which are orthogonal between the two metrics in a dynamic range of about 30 phon. This means it is not possible that the predictions of both metrics correlate highly with the subjective evaluations of overall loudness and therefore permit us to decide which of the two metrics is better.

2. METHOD

2.1 Participants

Eleven listeners, six of them females and five males, aged 21 to 30 years (median 23 years) participated in Experiment 1. 21 further listeners, 17 of them females and four males, aged 18 to 25 years (median 21 years) participated in Experiment 2. None of them participated in both experiments. All of them had normal hearing at audiometry frequencies from 125 Hz to 8 kHz, measured in octave steps and for each ear. They were students of psychology living in Germany.

2.2 Apparatus

The experiments were conducted in a double-walled sound-proof chamber manufactured by the Industrial Acoustics Company. The diotic stimuli with 48 kHz sampling rate and 24 bit resolution were D/A converted by a RME Hammerfall DSP Multiface II audio interface, led to a TDT HB7 amplifier and presented via Beyerdynamcis DT-48 headphones.

The free-field equalization for these headphones was implemented in software by simulating the passive network of [5] in Matlab.

2.3 Stimuli

All stimuli were recordings of real sounds, made with a Head acoustics binaural headset (BHS I) in Tokyo. They were converted to diotic sounds using its free-field transfer function. The sounds either consisted of a prominent portion and background noise themselves or they were constructed by adding two recordings. In the latter case, the prominent portion was one or several hammer hits, recorded in a sound-proof room, and the background noise was traffic noise, railway noise, helicopter noise or noise from a construction site. In these cases the added background noise was set to a rather soft loudness level of 58 phon.

Altogether, eight sounds were obtained that way. The largest difference between N5 and LL(P) was obtained for a single hammer hit with soft background noise, with LL(P) predicting 21 or 23 phon more, depending on the absolute level. The largest difference the other way round was obtained for a situation at a railway platform with a train horn as the prominent part. For this recording, the N5 is 8 to 9 phon higher than the LL(P). Further recordings inside a train or with several hammer hits provide intermediate conditions. These were chosen in a way to be roughly equidistant on the phon scale.

The eight sounds were played at two levels. In the first set of the eight sounds, the LL(P) was kept constant to 85 phon while the N5 varied between 64 and 93 phon. In the second set, the N5 was set constant to 68 phon while the LL(P) varied between 59 and 91 phon. Thus, 16 stimuli were obtained with subsets for which either N5 or LL(P) was constant and the other metric varied in a wide dynamic range between approximately 60 and 90 phon. The duration of each stimulus was 10 seconds with a Gaussian rise and fall time of 5 milliseconds.

2.4 Procedure

In both experiments, the stimuli were presented in six trials and in random order, resulting in 96 trials per participant. Loudness was judged using magnitude estimation by line length.

In Experiment 1, participants were asked to judge the overall loudness after a sound had finished. For that purpose a line appeared on the screen and the participant should adjust its length by moving the mouse so that it reflected his or her impression of overall loudness. In Experiment 2, the participants were additionally asked to continuously judge instantaneous loudness while the sound was played. After the sound, they also judged overall loudness. The theoretical maximum line length was 1260 pixels, leaving a margin of 10 pixels to each side of the monitor.

3. RESULTS

As the task for the subjects was a free magnitude estimation without a standard, the method chosen for averaging across the participants is the geometric mean. In the following, all data points are obtained by taking the geometric mean across participants and trials. Thus, each data point in Experiment 1 represents 66 trials, each data point in Experiment 2 represents 126 trials. Furthermore, calculated loudness will not be shown in the sone ratio scale but in the phon scale. For this reason, the subjective evaluations will also be logarithmized.

Figure 1 illustrates the results of Experiment 1. On the left hand side, the subjective evaluations are shown as a function of N5, on the right hand side as a function of LL(P). It can be seen that there is little variance in the conditions of constant LL(P), depicted by red circles and ranging from 384 to 570 pixels, which is a factor of 1.5. The conditions which should be constant according to N5 (blue squares), by contrast, range from 147 to 532 pixels, meaning a factor of 3.6.

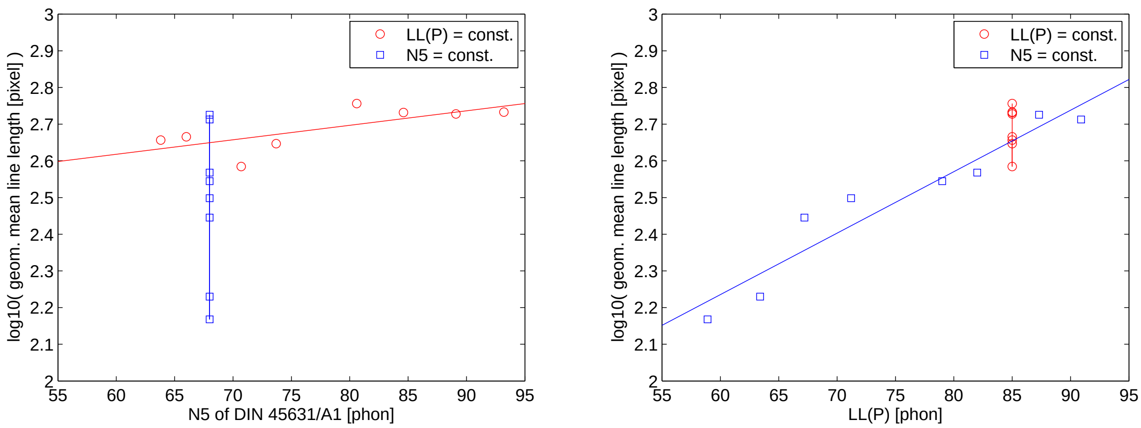

Figure 1 – Subjective evaluations of Experiment 1 as a function of N5 on the left hand side and LL(P) on the right hand side. Red circles indicate the set of conditions for constant LL(P), blue squares the set for constant N5. Regression lines are shown for each set of eight conditions.

The same analysis can be done for the subjective evaluations of overall loudness in Experiment 2, which is shown in Figure 2. The geometric mean subjective evaluations for the conditions of constant LL(P) range from 271 to 523 pixels, yielding a factor of 1.9. Those for the conditions of constant N5 range from 114 to 373 pixels, which is a factor of 3.3.

A concrete number to compare the performance of the two metrics as shown in Figures 1 and 2 is their correlation coefficient with the subjective evaluations. In Experiment 1, LL(P) and the subjective evaluations correlate with $r(14) = .944$, $p < .001$, while N5 and the subjective evaluations do not correlate significantly, $r(14) = .464$, $p = .07$. In Experiment 2, LL(P) and the subjective evaluations correlate with $r(14) = .901$, $p < .001$, N5 and the subjective evaluations show a correlation of $r(14) = .606$, $p <.05$.

The two experiments further allow us to test whether overall loudness was judged the same way independent of the fact that the participants did continuous judgment in Experiment 2 while they did nothing while listening to the sounds in Experiment 1. The correlation between the subjective evaluations in Experiment 1 and those of Experiment 2 is $r(14) = .963$, $p < .001$.

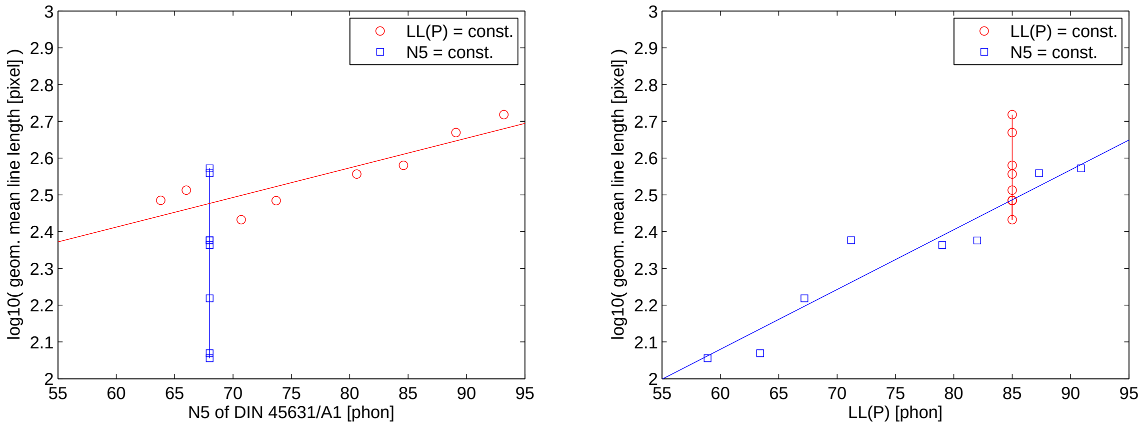

Figure 1 – Subjective evaluations of Experiment 2 as a function of N5 on the left hand side and LL(P) on the right hand side. Red circles indicate the set of conditions for constant LL(P), blue squares the set for constant N5. Regression lines are shown for each set of eight conditions.

4. CONCLUSIONS

A main difference of the present study towards older studies about the energy mean as a measure for overall loudness (see [6] for a review) is that it uses the LL(P) which is calculated out of loudness levels over time and not sound pressure levels over time. The second metric evaluated in this study, N5, is also obtained by a model which calculates loudness over time. Despite of calculating similar loudness for many sounds, their concepts are rather different.

Each of the present two experiments directly confronted N5 and LL(P). By using eight sounds which vary in a dynamic range of 30 phon for the one metric while their loudness is estimated as being constant by the other metric, conditions could be constructed which permit us an explicit comparison by looking at the correlation coefficients. The correlation between LL(P) and the subjective evaluations is higher than .90 in both experiments while it is just .46 and .60 for N5. This advantage of LL(P) can also be seen in Figures 1 and 2. The subjective evaluations of the constant N5 conditions vary in a wider range than they do for the constant LL(P) conditions. Altogether, the present results suggest a superiority of LL(P) for calculating overall loudness.

ACKNOWLEDGEMENTS

The authors thank ONO SOKKI for giving advice on the calculation program of DIN 45631/A1.

REFERENCES

1. DIN 45631/A1. Berechnung des Lautstärkepegels und der Lautheit aus dem Geräuschspektrum - Verfahren nach E. Zwicker - Änderung 1: Berechnung der Lautheit zeitvarianter Geräusche (Calculation of loudness level and loudness from the sound spectrum - Zwicker method - Amendment 1: Calculation of the loudness of time-variant sounds); 2010
2. Namba S, Kato T, Kuwano S. Evaluation of loudness level of time-varying sounds. Proc. Inter-noise 2011; Osaka, Japan, 2011
3. ISO 532. Acoustics - Method for calculating loudness level; 1975
4. Glasberg BF, Moore BCJ. A model of loudness applicable to time-varying sounds. Journal of the Audio Engineering Society 50, p. 331-342; 2002
5. Zwicker E, Maiwald D: Über das Freifeldübertragungsmaß des Kopfhörers DT 48 (On the Freefield response of the earphone DT 48). Acustica 13, p. 181-182; 1963
6. Kuwano S, Namba S. Loudness in the Laboratory, Part II: Non-Steady-State Sounds. In: Florentine M, Popper AN, Fay RR, editors. Springer Handbook of Auditory Research, Vol. 37, Loudness, p. 145-168. New York, Dordrecht, Heidelberg, London: Springer; 2011