Loudness experiments and models

1st semester projekt, DTU
Ørsted•DTU
Acoustic Technology
Fall 2007

Group 6
Troels Schmidt Lindgreen
Kristoffer Ahrens Dickow
Reynir Hilmisson
Title: Loudness experiments and models

Course: Acoustic communication

Project term: Fall semester 2007

Project group: 6

Participants: Troels Schmidt Lindgreen
Kristoffer Ahrens Dickow
Reynir Hilmisson

Supervisor: Torben Poulsen

Copies: 4

Pages: 17

Appendices: 1

Date of completion: 4th October 2007

Synopsis:

This report covers different sound perceptions within loudness experiments.

The experiments this report covers are:

1. Loudness adjustment
   The loudness of a probe tone has to be adjusted to half or twice the loudness of a reference tone.

2. Temporal loudness integration
   The loudness of a pair of tone pulses, differing in duration, has to be matched.

3. Spectral loudness summation
   The loudness of a pair of bandlimited noise burst, differing in bandwidth, has to be matched.

Test subjects are presented with audiometric headphones and a computer-controlled adaptive tracking procedure running in Matlab, which runs the actual psychoacoustic measurement.

The results support the theory of loudness but are influenced by the observer-expectancy effect.

No part of this report may be published in any form without the consent of the authors.
Contents

1 Loudness theory .......................... 1
  1.1 Loudness and loudness level ................. 1
  1.2 Temporal integration .......................... 2
  1.3 Spectral loudness summation ......................... 3
  1.4 Test procedure description ....................... 3

2 Observer-expectancy effect .................... 5
  2.1 Elimination of the observer-expectancy effect .............. 5
    2.1.1 The double-blind method ...................... 5
    2.1.2 Computer-controlled experiments .................. 6
  2.2 Unreliable results from group members .................. 6

3 Test results ............................. 7
  3.1 Loudness adjustment .......................... 7
    3.1.1 Discussion ................................... 7
  3.2 Temporal loudness integration ................... 10
    3.2.1 Discussion ................................... 10
  3.3 Spectral loudness summation ................... 11
    3.3.1 Loudness model predictions ................... 11
    3.3.2 Discussion ................................... 11
  3.4 Sources of error ................................ 12

4 Conclusion ................................ 15

Bibliography .................................. 17

A Loudness model predictions software ........... A1
Chapter 1. Loudness theory

1.1 Loudness and loudness level

The sone scale was established in order to avoid the mental confusion between sound pressure levels [dB SPL] and the subjective perception of loudness [sone]. Sone is a unit of perceived loudness $N$ after a proposal of S. Smith Stevens in 1936.

In acoustics, loudness is a subjective measure of the sound pressure. One sone is equivalent to 40 phon, which is defined as the equivalent loudness level $N_L$ of a 1 kHz tone at 40 dB SPL. The number of sones to a phon is defined so that a doubling of the number of sones sounds to the human ear like a doubling of the loudness. This corresponds to increasing the sound pressure level by approximately 10 dB, or increasing the mean square sound pressure by a factor of 10. [Poulsen, 2005a, p 27]

In figure 1.1, the y-axis is a logarithmic scale and the x-axis is linear in phone (dB SPL @ 1 kHz). Because dB is a logarithmic quantity the figure shows a double logarithmic relation. The straight part of the solid line corresponds to Steven’s power law, which in this case is [Poulsen, 2005a, p 28]:

$$N = kp^{0.6} \quad (1.1)$$

Where:
- $N$ is the loudness in sones
- $k = 0.01$
- $p$ is sound pressure in micropascal ($\mu$Pa)

Near the hearing threshold, the curve becomes steeper. This can be expressed in equation 1.1 by introducing the threshold sound pressure, $p_0$ [Poulsen, 2005a, p 28]:

$$N = k(p - p_0)^{0.6} \quad (1.2)$$

This is called a modified power function.

The straight part of the curve is given by [Poulsen, 2005a, p 28]:

$$N = 2^{L-40}/10 \quad (1.3)$$
5. Loudness

The term 'loudness' denotes the subjective perception of strength or powerfulness of a sound. The unit for loudness is Son or Sone. Note that 'loudness' and 'loudness level' are different concepts. Translation of terms:

- **Loudness**
  - **Unit:** Sone

- **Loudness Level**
  - **Unit:** Phone

Danish: Hørestyrke
German: Lautheit
French: Sonie

5.1 The loudness curve

The Sone scale was established in order to avoid the mental confusion between sound pressure levels (in dB) and the subjective perception of loudness: A 1 kHz tone at 80 dB SPL is not perceived double as loud as the same tone at 40 dB SPL. Figure 5-1 shows the relation between the Sone and the Phone scales. (Hint: for a 1 kHz tone, phone and dB SPL is the same number).

Arbitrarily it has been decided that one sone should correspond to 40 phones. The curve is based on a great number of loudness comparisons. The curve is called a loudness curve.

Figure 5-1: The loudness curve for a normal hearing test subject (solid line) and for a person with a cochlear hearing loss (dashed). From [Poulsen, 2005a, p 27].

Where:

- $N$ is the loudness [sone]
- $L$ is the loudness level [phon]

The curve shows that a doubling of the loudness corresponds to a 10 phone increase in loudness level, or a 10 dB increase in sound pressure level if we are dealing with a 1 kHz tone. In real life a rule of thumb says that a 10 dB increase is needed in order to perceive a doubling of the loudness.

The loudness curve becomes steeper near the hearing threshold. The steeper slope means that near the threshold the loudness increases rapidly for small changes in the sound level. This effect is called loudness recruitment [Poulsen, 2005a, p 28].

1.2 Temporal integration

The perception of loudness is related to both the intensity and duration of a sound. This means that short duration sounds (less than one second) are perceived lower than a sound with the same sound pressure level but longer duration. The growth of loudness as a function of duration is called temporal integration.

For example, a sound of constant intensity will be perceived to grow in loudness as 20, 50, 100, 200 ms samples are played up to a maximum of ~500 ms where the perception of loudness will stabilize, see figure 1.2.
The loudness growth can be described by an exponential relation [Poulsen, 2005a, p 30]:

\[ I(t) = I_\infty / (1 - e^{-t/\tau}) \]  \hspace{1cm} (1.4)

Where \( I(t) \) is the intensity of a signal of duration \( t \)
\( I_\infty \) is the intensity for a long duration
\( \tau \) is a time constant

Based on an extensive number of investigations, the best time constant is about 80 ms to 100 ms [Poulsen, 2005a, p 30]. Temporal integration has been measured by means of loudness comparisons of tone pulses of different durations. Figure 1.3 shows an example of such measurements. It can be seen that a 5 ms tone pulse needs a 17 dB higher level than a 640 ms tone (presented at 35 dB SPL) in order for the two pulses to have the same loudness.

1.3 Spectral loudness summation

The perception of loudness strongly depends on the frequency content of a sound. Loudness is not a simple function of physical intensity integrated across frequency. If this assumption would be true, loudness should be the same for sounds of different bandwidth but identical overall intensity.

In practice, loudness increases with bandwidth for sounds that have equal overall intensity. Thus, rather than by simple integration of intensity across frequency, loudness is better described by a spectral excitation pattern based quantity. This concept is based on the assumption that loudness may depend upon a summation of neural activity across different frequency channels [Ewert and Fobel, 2005, p 2].

1.4 Test procedure description

All experiments are done in Matlab, using a custom Matlab script, see [Ewert and Fobel, 2005] for procedure and equipment used.

All experiments uses an adaptive two-interval, two-alternative forced choice procedure. Two signal intervals are presented consecutively, with a pause interval. The subject tested has to determine which...
Temporal integration has been measured by means of loudness comparisons of tone pulses of different durations. Figure 5-4 shows an example of such measurements. It can be seen that a 5 ms tone pulse needs a 17 dB higher level than a 640 ms tone pulse (presented at 35 dB SPL) in order for the two pulses to have the same loudness. It can also be seen that the steepness of the curve is less at 95 dB compared to 35 dB.

Figure 1.3: Temporal integration, based on loudness comparisons of tone pulses of different durations. The lowest curve shows the threshold of the tone pulses.

interval has a particular signal feature or to answer a question after the presentation, depending on the experiment. In these tests, the test subject has to choose between two alternatives.

The presentations are varied differently in these experiments. The level of the presented signals where the interval with the louder stimulus is indicated, or two signals of the same level with different time intervals are presented. Also two signals varied with bandwidth are presented. If the test subject indicates the more intense signal as being louder, the adaptive procedure reduces the level of this signal by a certain amount (stepsize). Otherwise it would further increase the level using the same stepsize. This is called a one-up, one-down rule, and is used in all experiments this report.
Chapter 2

Observer-expectancy effect

Results of experiments can unconsciously be skewed by the observer/experimenter when the researcher expects a given result. The researcher unconsciously manipulates an experiment in order to find the wanted result. This is called the observer-expectancy effect [Wikipedia, 2007b].

The problem with the observer-expectancy effect can be significant – especially when testing human subjects. Robert Rosenthal has predicted that, when given the information to elementary school teachers that certain students are brighter than others, the teachers may unconsciously behave in ways that facilitate and encourage the ‘bright’ students’ success.

By this test Rosenthal concluded that biased expectancies can essentially affect reality and create self-fulfilling prophecies as a result. This is of cause an unwanted situation and actions to prevent that must be considered.

2.1 Elimination of the observer-expectancy effect

To eliminate the observer-expectancy effect, double-blind methodology is often used.

2.1.1 The double-blind method

The double blind method prevents research outcomes from being influenced by the possible observer bias from the observer-expectancy effect by lessen the influence of the prejudices and unintentional physical clues on the results. [Wikipedia, 2007a]

In contrast to an open experiment, the blinded experiments hides a factor from the test subject or the researcher. In double blinded experiment, neither the individuals nor the researchers know who belongs to the control group or the experimental group. The method relies on a third party to assign the subjects to the experiment. The key that identifies the subject is kept by the third party until the series of experiments is over. First after the experiment is finished and all the data are recorded the researchers learn which individuals are which. [Wikipedia, 2007a]
2.1.2 Computer-controlled experiments

Computer-controlled experiments can be double-blind experiments, since software should not cause any bias. In this case the the part of the software that provides interaction with the human is the blinded researcher, while the part of the software that defines the key is the third party. An example is the ABX test used in this report, where the human subject has to identify an unknown stimulus X as being either A or B.

2.2 Unreliable results from group members

The experiments covered in this report are computer-controlled and hereby double-blinded. But the fact that the test subjects and the researchers are the same persons, the test subjects have a bias knowledge hence the results are unreliable in some form.

The results are not in any form useless, but have been influenced by prior knowledge to what the result should be in general. The fact that the software used in the experiments uses an adaptive two-interval, two-alternative forced choice procedure the influence from biased knowledge is reduced, but not prevented completely. Especially when choosing which sound has the greatest loudness between a pulse of 10 ms and a pulse of 160 ms the knowledge can change the result considerably. Because the group and thereby the test subjects knew that the perception of loudness needs some time to build up, unintentionally the group will favor the short puls durations over the long pulses due to the observer-expectancy effect.
Chapter 3

Test results

The experimental results in this report are all based on the methods described in the exercise guide, [Ewert and Fobel, 2005]. The test subjects are the three group members. Each test subject has been tested twice in each of the three experiments – two runs are included in all the test procedures running in Matlab, such that when a test subject has completed all the steps in a given procedure he has been tested twice. Results are based on mean values of the two test results for each test subject.

3.1 Loudness adjustment

The test results of this procedure running in Matlab are shown in figure 3.1. The mean value is plotted as well as the individual results. Notice that the x-axis is logarithmic.

Since the test only considers two loudness levels, e.g. half the loudness and twice the loudness, it cannot be concluded directly from looking at the results whether the x-axis should be linear or logarithmic, but in order to do any kind of graphical comparison to the sone scale and Steven’s Power Law a logarithmic axis is needed.

The points are connected with a straight line, which according to Steven’s Power Law would be correct, but it must be emphasized that this relation is not shown by the results – the line is added because the relation is known from theory, and because it makes the figure more readable.

3.1.1 Discussion

As seen in figure 3.1 the results from the different test subjects vary about 8 dB at both half and twice the loudness of the reference. At half the loudness the results of the measurements are in the approximate range from 49 dB - 57 dB, and at twice the loudness the range is approximately 67 dB - 75 dB. The lines meet at around 'Equal loudness' which again corresponds to ∼60 dB.

The results is of course what is expected since the reference stimulus is exactly 60 dB SPL. However this shows that the test subject who perceived half the loudness to be around 49 dB is also the one who perceived twice the loudness to be 75 dB. The same thing goes for the other test subjects – the greater the difference (relative to the reference stimulus) is, at half the loudness, the greater is the difference at twice the loudness.

From this it can be concluded that the test subjects perceive loudness differently from each other, but that the results still show consistancy with the logarithmic scale because equal loudness level is at the same level.
Figure 3.1: Results from the loudness adjustment experiment.

as the reference stimulus. Basically this is as expected from theory, and it justifies connecting the points with a straight line on the double logarithmic coordinate system.

The mean, shown in figure 3.1 is compared to the expected result which is derived from Steven’s Power Law. The expected values are ±10 dB compared to the reference stimulus because Steven’s exponent for loudness is 0.3 and it can be shown from [Poulsen, 2005b, p 12] that every time 10 dB is added to the stimulus ($S_{dB}$), the perceived loudness ($N$) doubles.

$$N = \left(\frac{S}{S_0}\right)^{0.3}$$

$$\frac{1}{0.3} \cdot 10 \log (N) = 10 \log \left(\frac{S}{S_0}\right) = S_{dB} \tag{3.1}$$

Thus with 10 dB added:

$$10 + S_{dB} = \frac{1}{0.3} \cdot 10 \log (N) + 10$$

$$= \frac{1}{0.3} \cdot 10 \log (N) + \frac{1}{0.3} \cdot 0.3 \cdot 10 \log (10)$$

$$= \frac{1}{0.3} \cdot 10 \log (N) + \frac{1}{0.3} \cdot 10 \log (10^{0.3})$$

$$= \frac{1}{0.3} \cdot 10 \log (N \cdot 10^{0.3})$$

$$10 + S_{dB} = \frac{1}{0.3} \cdot 10 \log (2N) \tag{3.2}$$

Equation (3.1) and (3.2) shows that in general a 10 dB increase in stimulus is needed in order to perceive a doubling in loudness. In figure 3.2 a comparison between the mean of the test results and the expected result of Steven’s Power Law is shown.

From [13], page [1] the straight part of the sone scale in figure [1.1] is given by $N = 2^{\frac{S_{dB}}{10}}$. This corresponds
to Steven’s Power Law which is shown with the mean of the test result in figure 3.2. The measurements are not far from the expected results, considering that only three test subjects were involved in the experiment. From the relation $L \sim I^\alpha$ (See [Ewert and Fobel, 2005, p 4]) the exponent $\alpha$ is calculated by determining the values of the mean by zooming in on the graph in Matlab. These are listed in table 3.1.

$$\begin{array}{|c|c|}
\hline
\text{Relative loudness level} & \text{Relative dB-level} \\
\hline
\text{Twice} & 14.25 \\
\text{Half} & -11.14 \\
\text{Twice} & 6.42 \\
\text{Half} & -3.05 \\
\hline
\end{array}$$

Table 3.1: Mean values from loudness adjustment experiment.

The average is $+10.33$ dB relative to the reference for twice the loudness and $-7.095$ dB (relative) for half the loudness. These results are now used to estimate $\alpha$.

$\alpha$ is the inverse slope of the mean line in figure 3.2 since the stimulus level is on the y-axis and the perceived loudness on the x-axis. Which gives:

$$\alpha = \frac{10 \log 2 - 10 \log \frac{1}{2}}{10.33 - (-7.095)} = \frac{10 \log 4}{17.425} = 0.346$$

(3.3)

This is relatively close to Steven’s exponent, $e_{\text{loudness@1kHz}} = 0.3$.

For most people it is a difficult task to evaluate relative loudness. This means that a rather big variance in the test results is expected. If the exponent were to be calculated in the extremes of the mean $\alpha$ would be ranging from $\frac{10 \log 4}{14.25 - (-7.095)} = 0.237$ to $\frac{10 \log 4}{6.42 - (-3.05)} = 0.636$. 

Figure 3.2: Mean of loudness adjustment test results compared to expected results.
3.2 Temporal loudness integration

Results from this procedure running in Matlab are shown in figure 3.3. The mean value is plotted as well as the individual results. Notice that the x-axis is logarithmic. For both the mean value and the individual results the variances are indicated as vertical lines centered around the calculated result.

Figure 3.3: Results from the temporal loudness integration experiment.

3.2.1 Discussion

Figure 3.3 shows a variance at 10 ms, and also considerable variance for the short duration pulses (duration < 160 ms), whereas at 160 ms pulse duration the variance is small. At 640 ms the variance grows a bit again. This is as expected since the reference tone has a duration of 160 ms, and the uncertainty should be small when comparing two tone pulses of the same duration. At 80 ms the test subject ‘RH’ has an unexpected peak. The reason for this is, that the subject at some point confused ‘up’ and ‘down’ in the test procedure and did not realize the mistake until the test tone was turned up quite loud. However, it was decided to keep the results rather than having the test subject do the test again, since the results were known by the test subject, and the results of a second test could therefore be influenced by the expectations of the test subject accordingly to the observer-expectancy effect, see chapter 2.

In order to find the time constant, $I_\infty$ must be estimated. Since the reference tone in the experiment has a duration of 160 ms it can not be assumed that the reference level of 55 dB corresponds to the intensity for a long duration. In the following calculations $I_\infty$ is set to 54 dB. This is an estimate based on experimental results from figure 3.3.

The relation between the intensity of a signal of duration $t$ and a signal with a long duration is [Poulsen, 2005a, p 30]:
October 2007  

Chapter 3. Test results

\[ I_{\text{in}}(t) = 10 \log \left( \frac{10^{\frac{I_\infty}{10}}}{1 - e^{-\frac{t}{\tau}}} \right), \]  

(3.4)

Where:

\( \tau \) is the time constant.

Graphical comparisons between equation (3.4) and the test results in figure 3.3 are given in figure 3.4.

Based on figure 3.4 the results of the experiment points towards a time constant of \(~50\) ms compared to the expected range from 80 ms to 100 ms according to [Poulsen, 2005a, p 30]. This is seen by the curves for \( \tau = 40 \) ms and \( \tau = 60 \) ms, which approximates the experiment results.

3.3 Spectral loudness summation

The output of Matlab after running the spectral loudness experiment on three test subjects is shown in figure 3.5. Mean values are plotted as well as individual test results.

3.3.1 Loudness model predictions

As described in [Ewert and Fobel, 2005, p 5] certain prediction models for spectral loudness summation, can be used. These are implemented on a Matlab script at the test computer. The results is shown in figure 3.6. See appendix A for the script code and a calculation example. The results are based on the Moore and Glasberg model and the DIN 45631 model comparing to a reference value based on 20 runs of the given model. The mean value of the group’s test results is plotted for comparison.

3.3.2 Discussion

In figure 3.5 measurements in the narrow (200 Hz - 800 Hz) bands show large variance. At broader bands the variance decreases which indicates that the subjects tend to have more consistency in their answers. In some narrow bands some perceived the signals as 10 dB lower than others. If this is compared to expected results, predicted by models as in figure 3.6 the measured result indicates that the test subjects in this particular group perceives narrow band noise louder than an average normal hearing person. However, as it
Chapter 3. Test results

Figure 3.5: Results of spectral loudness summation experiment

will be discussed in the next section, this can be caused by the fact that the test subjects knew beforehand that narrow band noise would be perceived softer than broadband noise with the same intensity accordingly to the observer-expectancy effect, see chapter 2.

3.4 Sources of error

The results above vary from the expected ones based on the theory of loudness. The main reason for this is the small number of test subjects. The theory of loudness is made by multiple experiments with many test subjects so that a mean can be calculated for different types of experiments. These means are what the theory is based on, and are also the values the results from this report are compared to. Three test subjects is not enough to calculate a representative mean to compare with the theory because the weight of one abnormal result will have a big impact in the mean for the test subjects. In a large group of test subjects such an abnormal result will be insignificant for the mean and thereby have no influence.

The experiments were conducted in a noisy room. There was a lot of talk all around the test subject and concentrating was difficult after hours of lectures and calculation. All these factors combined caused a discontinuous noise, disturbing the test results. The discontinuocy of the noise entails, the noise influenced some signal pulse series, while not others. This had a major impact on the results, especially when comparing the loudness of a faint click with a faint pulse of longer duration with the background noise being very loud at moments. This is easily seen on figure 3.7.

When measuring temporal loudness integration in a noisy room, it is hard for the test subject not to be influenced by the observer-expectancy effect, knowing that short interval tones are actually louder than they sound (see figure 1.7), so the brain tries to cheat the test by choosing the click over the long pulse when the pulses sounds equally loud.
The observer-expectancy effect is easily seen when the test subject did the same test a second time well knowing the result of the first experiment was not equal to the theory. This can be seen on figure 3.6.

Figure 3.6: Spectral loudness summation. Mean of group results compared to loudness model predictions.
Chapter 3. Test results

Figure 3.7: Test results from the loudness temporal integration experiment. 'TL' was measured in a room with silence. 'TL2' was measured in a noisy room with discontinuous loudness of noise.

Figure 3.8: RH show the first result of the experiment. RH2 shows the result of another test after viewing the result of the first and knowing Stevens' Powerlaw. It is clearly seen the test subject is trying to equalize the difference between the first result and Steven’s Powerlaw. In fact, the next result becomes steeper than Steven’s Powerlaw.
Conclusion

This report shows and concludes different experiments and methods regarding loudness. The results shows that the human sound perception varies from person to person. Not only is the human ear sensitive to sounds that differ in time duration but also sounds containing different frequencies. Loudness, which is a wide concept, is essential to get an understanding of the human auditory system. Main conclusions from specific tests in this report are:

- Stevens Power Law is a good approximation to the perception of loudness.
- Sounds of duration $> 500$ ms are perceived as being louder than sounds of shorter duration with the same overall intensity.
- Broadband noise is perceived louder than narrowband noise with the same overall intensity.
Bibliography


Loudness model predictions software

This appendix describes the Matlab code used to make figure 3.6.

First the initial conditions are declared.

```matlab
% Initial matlab codes.
clear
format compact
% Beginning of the program.
Hz = 6400;   % With of bandwidth to be compared with the reference
            % bandwidth of 3200 Hz @ 55 dB SPL
dB = 55;     % Initial guess of SPL for the given bandwidth to be equal to
            % the reference.
res = 0.25;  % Declaration of variable. 'res' is the resolution in dB
model = 'm'; % 'd' for DIN model or 'm' for Moore model.
```

A mean over 20 calculations of the reference value is used as the final reference value. The reason for taking the mean value is that fact that the model gives a slightly different result each time it is run. This is due to the fact that both the DIN 45631 model and the Moore & Glasberg model uses a finite length noise sample to calibrate the data. Noise have a distinct frequency response but only in average. Not at a particular time. Because a finite length noise is used the frequency response will not be exactly the same every time. If, however, it was possible to use an infinite length noise sample, the result should be exactly the same every time – that being in theory.

```matlab
% Calculation of reference value.
ref = 0;     % Declaration of variable. 'ref' is used to find the mean of the reference.
fprintf('Calculating mean reference value:\n');
for n=1:20
    LoudSumModel(55,3200,model);
    ref=(ans+ref);
end
refvalue=ref/n;
```
To calculate the number of dB SPL a given bandwidth must be amplified to sound like the reference value, two if sentences are used. If the given value is smaller than the reference value the sound pressure level is increased by the resolution value until the reference value is equal (within uncertainty in the resolution) to the value for the given test bandwidth.

Finally the result is plotted:

For an example of a calculation using the Matlab script, see the next page.
Example of calculation for the bandwith 6400 Hz, beginning calculations at 50 dB SPL with a resolution of 0.25 dB SPL using the Moore & Glasberg model:

```
Calculating mean reference value:
7.78 sone
7.81 sone
7.85 sone
7.78 sone
7.79 sone
7.77 sone
7.82 sone
7.87 sone
7.83 sone
7.83 sone
7.80 sone
7.76 sone
7.81 sone
7.84 sone
7.74 sone
7.82 sone
7.83 sone
7.75 sone
7.88 sone
7.78 sone

Mean reference value is 7.807629e+000 sone.

Calculating SPL amplification for 6400 Hz:
7.84 sone
7.77 sone

The bandwith 6400 Hz has to be amplified 4.975000e+001 dB to match the bandwith of 3200 Hz @ 55 dB SPL.
```