Contents

1. General Introduction

In everyday life the term loudness is not clearly defined. Often it is a mixture of different hearing sensations. In psychoacoustics the sensation loudness corresponds to the physical property of intensity of the sound.

However, the sensation loudness depends not only on the intensity of the sound, but also varies with other physical parameters such as e.g. frequency-spectrum and duration. One way to measure loudness is to ask the subjects to adjust the level of one signal to get the same hearing impression of loudness as a reference signal (loudness matching procedure). As the reference signal usually a 1 kHz tone is used. With this procedure curves of equal loudness impression can be obtained as a function of one (or several) physical parameters. When loudness is measured as a function of frequency this curves are called equal-loudness level contours ($ELLC$) or isophons. The unit of the loudness is phon and is defined as the intensity of a 1 kHz tone in dB SPL producing the same loudness sensation as the signal at a given level.

The shape of the curves, shown in Fig. 1.1 reflects the different loudness impressions when reproducing the same sound at different levels. A minimum is observed at 4 kHz for the lowest $ELLC$. This reflects the high sensitivity of the ear to frequencies around 4 kHz. The increase of the curves on both sides indicate that at very low level the ear is relatively insensitive for very low and high frequencies. In everyday life the phenomenon is known when listening to music at a soft level over a hi-fi system. The hearing impression is 'thin' with a dominant mid frequency range. To compensate for this phenomenon some audio amplifiers have a 'loudness' function which raises the low and high frequencies. But as one can see from Fig. 1.1 at higher levels the curves are compressed. Since the loudness function of most amplifiers is independent of the level, at high levels, where the ear is almost equal sensitive for all frequencies, the sound is described as 'boomy'.

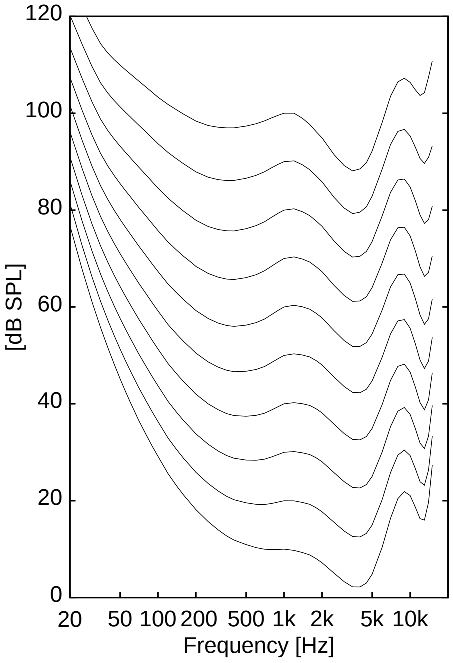

Figure 1.1.: *Equal-loudness level contours defined in ISO 226 in the range of 10 to 100 phon in steps of 10 phon. The ordinate represents the sound pressure level.*

Another application of the equal-loudness level contours is the measurement of loudness level, where a frequency weighting is used to obtain a better agreement between measured and subjectively perceived loudness impression. The weighting reflects the frequency dependent sensitivity of the ear. Different weighting curves are defined in DIN/IEC 651 [1981]. The A-weighting reflects the inverse 30 phon contour. The B-weighting corresponds to the curves between 30 and 60 phon and the C-weighting corresponds to higher *ELLC*.

Thus, equal-loudness level contours are of fundamental interest for the loudness calculation and therefore reliable data are important. New data for the equal-loudness level contours show considerable differences in the results of different laboratories, although the investigations were performed using the same method. All new results from Betke [1991], Fastl et al. [1990], Suzuki et al. [1989], Watanabe & Møller [1990] show the uniform tendency to lie above the standard values of ISO 226 [1961]. A detailed overview over former data and the reason for the re-determination is given in the second chapter of this study.

The new data cannot directly be used for the definition of a new standard since there are considerably differences between the different studies. A prominent example are the two newer measurements from Betke [1991] and Fastl et al. [1990] which were carried out under well defined and comparable conditions. For a frequency of 125 Hz data from Betke are up to 11 dB higher than data from Fastl et al.

A systematic investigation and comparison of these two works was done by Gabriel [1996]. Differences between both studies were examined in detail and several parameters of the measurement setup were systematically varied to determine the influence on the results. Gabriel showed that the largest differences between the two studies could be attributed to the employed level range which was presented to the subjects. The judgments of the subjects were mostly influenced by the absolute position of the level range of the presented test-tones. This effect increased with increasing frequency distance between test- and reference-tone. This so-called range-effect was firstly systematically investigated by Gabriel for the determination of $ELLC$. Gabriel proposed a simple adaptation model which explained the effect in a first approximation.

In the present work the model was also applied to data from the literature. The model was explored in detail and its requirements and limitations were discussed in Chapter 2. Furthermore a measurement procedure considering the findings of Gabriel was proposed.

An adaptive procedure was used for the determination of the equal-loudness level contours. This procedure is in agreement with the two alternative forced choice procedure recommended by (ISO/TC43/WG1/N122 [1988]).

The results are compared to the one obtained with the method of constant stimuli in Chapter 3.

In Chapter 4 the range effect was investigated for equal-loudness levels measured with an adaptive procedure. Two different experimental parameters – starting level and initial step-size – were systematically varied to quantify their influence on the obtained results and to estimate the maximum measurement error in the adaptive procedure. In the literature it is argued that the influence of the employed starting level is minimized when using an adaptive procedure with interleaved tracks. Thus, an interleaved adaptive procedure was tested which is comparable to the method proposed by Florentine et al. [1996] to measure equal-loudness level contours. First a simple procedure with two tracks similar to Jesteadt [1980] and then an interleaved procedure with four tracks was tested. This interleaved procedures are described in Chapter 5.

In addition, the influence of interindividual differences in absolute threshold of hearing on equal-loudness level contours was investigated in this study.

2. Analysis of data from the literature concerning equal-loudness level contours

Several laboratories published data for equal-loudness level contours. The first systematic measured data were obtained by Kingsburry [1927], followed by Fletcher & Munson [1933], Churcher & King [1937], Zwicker & Feldkeller [1955] and Robinson & Dadson [1956]. In 1961 data were standardized in ISO 226 [1961] which was mainly based on data from Robinson & Dadson [1956]. In Figure 1.1 equal-loudness level contours ($ELLC$) defined as in ISO 226 are plotted from 10 up-to 100 phon in 10 phon steps. As one can see from Fig. 1.1, $ELLC$ in the frequency range around 400 Hz are lower compared to those around 1 kHz. Hence, the human ear should be more sensitive to a 400 Hz tone than to a 1 kHz tone. However, loudness comparison measurements from Fastl and Zwicker [1987] in this frequency range yielded to different results. Because of these discrepancies a revision of ISO 226 was decided. Newer results, e.g. Betke [1991], Fastl et al. [1990], Gabriel [1996], Suzuki et al. [1989] and Watanabe & Møller [1990] show the tendency to higher values than the standard values of ISO 226 [1961] for all frequencies below 1 kHz.

However, the new data cannot be used as a new base of standardization, since they show considerable differences among different laboratories. Even within the same working group, different values are published for the same *ELLC* in the literature. A comparison between the paper from Møller & Andresen [1984] and Watanabe & Møller [1990] is given in Appendix A.1.

Differences between newer data are remarkable because all results were obtained under the same recommended measurement procedure ISO/TC43/WG1/N122 [1988] which is a two-alternative forced choice procedure (2-AFC). However, the interpretation and application of the 2-AFC procedure was different in these studies. Whereas most research groups used the method of constant stimuli, Watanabe & Møller [1990] employed an adaptive procedure. But also for the studies using the method of constant stimuli, results differed markedly.

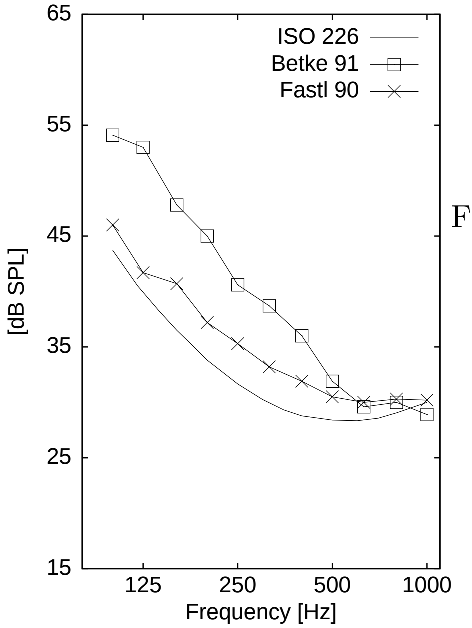

Figure 2.1.: *30 phon curves from Betke (□), Fastl (×) and data from ISO 226 (solid line) in the frequency range from 100 to 1000 Hz.*

For example, in Fig. 2.1 two test-series from Betke [1991] and Fastl et al. [1990] are shown. Both curves are higher than data from ISO226. Data from Betke lie above the results from Fastl. For a frequency of 125 Hz loudness level from Betke [1991] is about 11 dB higher than the value from Fastl et al. [1990]. Deviations between former and recent results were expected. They could be attributed to the employed measurement devices (e.g. loudspeakers, amplifiers). However, this is not the case for the new data.Thus, deviations between data from the literature from different studies should be contributed to differences in the experimental setup. For example, the level range of test-tones in the 30 phon-experiment of Betke was 32 dB whereas Fastl employed a range of only 16 dB. In addition, the mean level calculated over the presented level was higher in Betke's experiments. Betke tried to minimize the influence of the level range by presenting a large level range. The influence of the level range was investigated by Garner [1953]. In his experiments he reported about the tendency that the loudness level was close to the midpoint of the presented comparison tones. The results were strongly influenced by the employed range of the presented stimuli.

In the work of Gabriel [1996] the reason for the differences between the results from Betke [1991] and Fastl et al. [1990] were systematically investigated. The influence of the measurement room, the individual subject, the choice of test-tone levels, the number of test-tone level presentations, the size of the level range and the absolute position of the level range were examined. In agreement with Garner [1953], the experiments of Gabriel indicated that the absolute position of the level range was the main reason for the deviation between the data, whereas other parameters had minor influence on the equal-loudness level contours.

The influence of the size and the absolute position of the level range was frequency dependent and increased for large frequency distances between test- and reference-tone. The effect was called 'range-effect' and was explained in a first approximation by a simple adaptation model. In the following requirements and limitations of the model will be discussed.

2.1. Model for the range effect proposed by Gabriel

The aim of the model proposed by Gabriel was to predict the amount of the range effect, depending on the employed measurement configuration. Applied to data from the literature, deviations between different laboratories should be minimized.

The concept of the model for the explanation of the range effect was based on a model proposed by Helson [1964]. He proposed a model to calculate for absolute loudness judgements by accounting for the weighted mean intensity of all stimuli affecting the judgement.

$$logAL = (1 - w)logS + wlogC \tag{2.1}$$

AL = adaptation level
S = geometric mean intensity of all stimuli
C = intensity of the context stimulus
w = relative weighting of the two components

A comparable model was proposed by Marks and Warner [1991] who performed loudness experiments using a magnitude estimation procedure. In their experiments the influence of the employed level range on the resulting loudness was also shown.

Transferred to the paradigm of loudness comparisons the point of subjective equality (PSE) would not only be determined by an 'internal' absolute value (ABS), but could also be influenced by a relative value (REL). This value REL which could be derived from the level range of the presented test-tones. The extent of the influence was proportional to the 'internal' absolute value and its distance to the relative value with a frequency dependent weighting-factor w. This weighting factor reflected the effect that results were strongest biased for large frequencies distances between reference- and test-tone (see Gabriel [1996]).

In the work from Gabriel the following model was proposed, whereby in the first approximation REL was assumed to correspond to the mean level (ML) of the employed test-tone level range:

$$PSE = ABS + w(REL - ABS) \tag{2.2}$$

Evidently, this concept requires the existence of the absolute value ABS. Transferred to the loudness comparison this means the acceptance of the possibility of an absolute comparison judgment. If ABS is comparable between individuals, reproducible values should be obtained. Only in this case a standardization of equal-loudness level contours would make sense. ABS can be calculated as:

$$ABS = \frac{1}{1-w} PSE - \frac{w}{1-w} REL \tag{2.3}$$

The PSE-judgement submitted by the subject emerges from the ABS-measure through a change in the direction of REL (Eq. 2.2), which is determined by the presented test-tones. The value of w must be less than 1 (for a $w = 1$, the value ABS would be insignificant). The value REL reflects the experimental design. Gabriel has determined that REL depends on the absolute position of level range of the test-tones, presented to the subject. Results could be modeled, if REL is calculated from the presented maximum- and minimum level. In the determination of REL Gabriel accounted for the stronger influence of the maximum level (than the influence of the minimum level), by the following empirically determined equation: [1]

$$REL = 0.8\ MAX + 0.2\ MIN \tag{2.4}$$

The shape of w was chsoen to be similar to the shape of an auditory filter as proposed by Patterson et al. [1974] on the basis of notched-noise data.

---

[1] In the literature it is common to calculate percentile values from the distribution of the employed level. In this example percentile values are calculated by taking only the max and min value into account. This could be done for Gabriel's data since they were equally distributed on a level scale [in dB]. Thus Eq. 2.4 corresponds to the L_{20}.

For loudness comparisons w reflects that the influence of the experimental design increases for larger frequency distances between test- and reference-tone. The curve calculated from Eq. 2.5 is shown in Fig. 2.2.

$$w = -0.1 * 10log((1-r)*(1+p*g)*exp(-p*g)+r) \qquad (2.5)$$

$r = 0,00001$
$p = 46.4$
$w = PSE/ML$
$ML = \sum_{i=1}^{N} T_i/N$
T_i= level of the presented test-tones
N= number of trials
$g = |(f_{TT} - f_{RT}|/f_{RT}$
f_{TT} = frequency of the test-tone
f_{RT} = frequency of the reference-tone

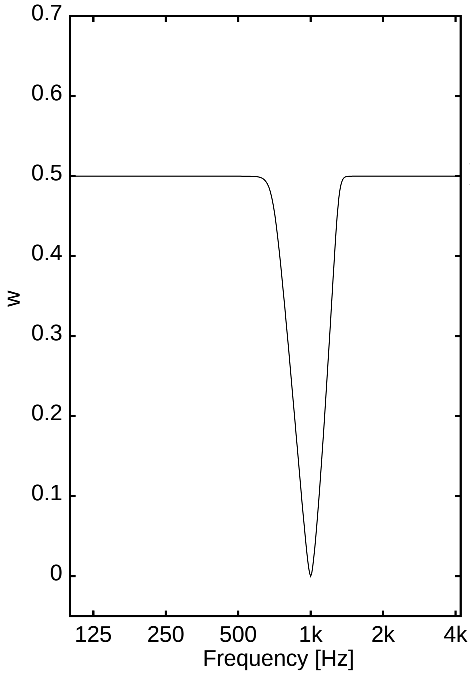

Figure 2.2.: *Weighting function for modeling the range-effect proposed by Gabriel. The size of w reflects the frequency dependent amount of the range effect.*

However, a systematic investigation of the model assumption is still lacking. Gabriel indicated that a level measure may not be optimal for the

determination of REL from the distribution of the test-signals. The higher weighting of higher levels suggests the application of loudness levels instead of levels in [dB SPL] with Eq. 2.4.
In the next section the model will be applied to loudness data published from different laboratories, to investigate if differences between $ELLC$ data could be minimized.

2.2. Application of the model to data from the literature

Unfortunately in most studies in the literature not all measurement parameters were mentioned which are necessary for the calculation of ABS (see Eq. 2.3) . Therefore, in this section only test-series were considered, where the test-design was described in detail. The model was first applied with the parameters proposed by Gabriel to four test-series to test if differences between the results of different laboratories can be minimized. In addition, to quantify the influence of the different parameters on the results and to look if an improvement (in the sense of minimizing differences between loudness data from the literature) could be achieved, the model prediction were calculated with different realization of REL and w.

In Fig. 2.3 (left panel) test-series from Betke [1991] and Fastl et al. [1990] are shown. In addition data from Gabriel [1996] which were obtained under the same experimental conditions as in the studies from Betke (denoted as Betke-BG) and from Fastl (denoted as Fastl-BG) are shown. Data from Betke lie above the results from Fastl. This is true for the original data as well as for the reproduced measurements from Gabriel. These differences increase for large frequency distances between test- and reference-tone. All four test series lie above ISO 226, close to that curve are the reproduced measurements from Gabriel under the condition from Betke, the highest values are the original data from Betke. The next lower curve indicates the

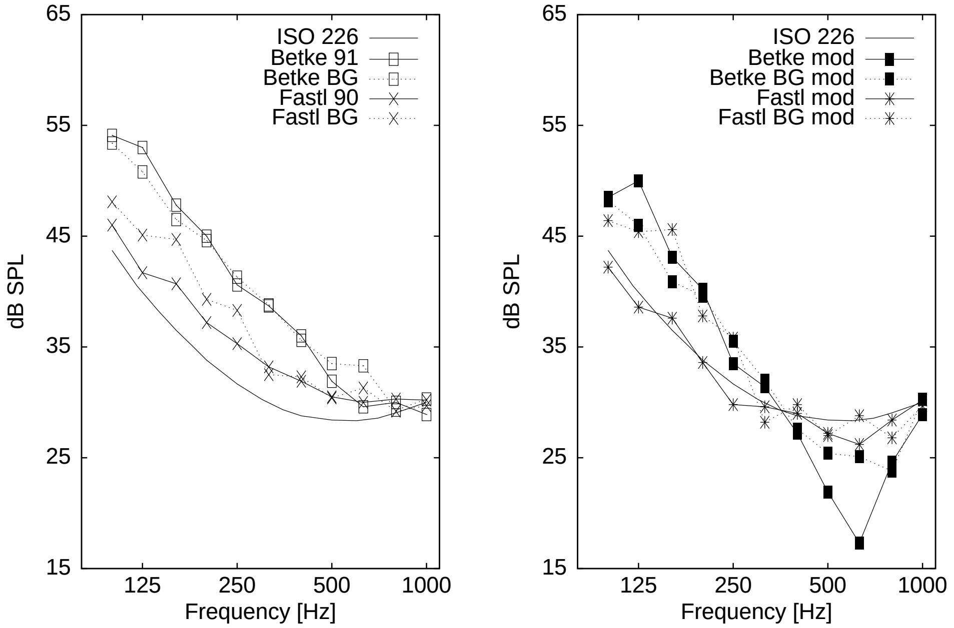

Figure 2.3.: *Left: 30- phon curves from Betke (□) and Fastl (×) and reproduced data from Gabriel (dotted lines) Right:* Calculated ABS values using the values of the left figure with Eq. 2.2. REL was calculated as L_{20} from Eq. 2.4.

reproduced data from Gabriel determined with the condition from Fastl and the lowest contour (except for the data of ISO 226) are the original data from Fastl. Differences between the sets of Betke and Fastl, as well for the reproduced data, increase for lower test-tone frequencies. For most of the frequencies the reproduced data lie close to the original data, but at 160 Hz reproduced data from Gabriel are 4 dB higher than the original data from Fastl. At 630 Hz reproduced data from Gabriel (Betke-BG) are 3.7 dB higher than the original data from Betke, this difference is by a factor 2 larger than the difference between the original data from Betke and Fastl.

The model was applied to these data to test if differences between the studies could be explained by accounting for the range effect. Under the assumption of a frequency dependent weighting-factor w, shown in Fig. 2.2, ABS values from the measurements (Fig. 2.3, left panel) were calculated and shown in the right panel of Fig. 2.3. The relative value REL (see Eq. 2.4), was calculated on the base of 80 % max and 20 % min level of the presented test-signals at the respective frequency. In general, the results are not convincing. Some measurement differences are lowered according to the model's objectives Eq. 2.2. But on average the variance between the data increases. Especially for the original data from Betke, the model predicts unrealistic low values for frequencies in the frequency range from 500 to 800 Hz. For example, at 630 Hz the resulting ABS value is reduced to 17.3 dB. This is probably well below a realistic value. In this case the influence of the measured PSE is to strong. Gabriel's original data for 630 Hz are 3.7 dB higher than Betke's original data at this frequency. The model increases this difference to 7.8 dB.

To show how a small change in the measured REL data is enlarged in the resulting ABS due to the model, different ABS values were calculated for some possible values of REL. The results are shown in Table 2.1. At 630 Hz the weighting function is close to 0.5 (see Eq. 2.5 and Fig. 2.2). In addition ABS values are calculated for a w of 0.3 which is the weighting factor for 800 Hz. With this smaller value w, unreasonable low ABS values could be avoided. Because the weighting function has a relatively sharp minimum in this region and is only calculated for a small amount of frequencies by Gabriel this change in the weighting factor is not unreasonable. However, even with the smaller w the obtained results are still below all measured data. In section 2.3 a systematically analysis will be done for the parameters of the model.

The consequence of the model given by Eq. 2.2 becomes clearer when data of the four single test-series are averaged. Fig. 2.4 shows mean and standard deviation of the original data on the left and the predicted ABS data on the right panel. Evidently the model (Eq. 2.2) reduces at some frequencies the deviations between different experimental designs (frequencies be-

PSE	ABS (w=0.5)	ABS (w=0.3)
27	12.5	20
30	18.5	24
33	24.5	29

Table 2.1.: *Calculated ABS values for three different values of REL and two values for w.*

low 400 Hz). However, at higher frequencies and especially at 630 Hz the contrary effect occurs. Eq. 2.2 shifts the contour at low frequencies to lower values, here of about 3 dB below 200 Hz reflecting the experimental procedure. More levels below the resulting PSE were offered at low frequencies.

The value REL was just estimated empirically by Gabriel, a different determination of REL will be proposed in section 2.3.1.

Application of the model to higher and lower level

Although the model could not minimize the discrepancies between Betke and Fastl data it may reduce the differences between the new data from the studies of (Betke [1991], Fastl et al. [1990], Suzuki et al. [1989] and Watanabe & Møller [1990]), to a reasonable degree. An overview over the employed procedures and the measured contours is given in Table 2.2.

In contrast to the last section, data were calculated with the model for others than the 30 phon contour, in order to see if the modeled data show the global trend of moving closer together.

The analysis was restricted to the frequency range below 1 kHz, where most data are available and the largest discrepancies occur. The REL was calculated with Eq. 2.4 proposed by Gabriel [1996]. Since in the literature only little information about the presented levels are available, a calculation regarding the distribution of the employed level was not possible.

The original loudness data and the estimated ABS data are shown in the

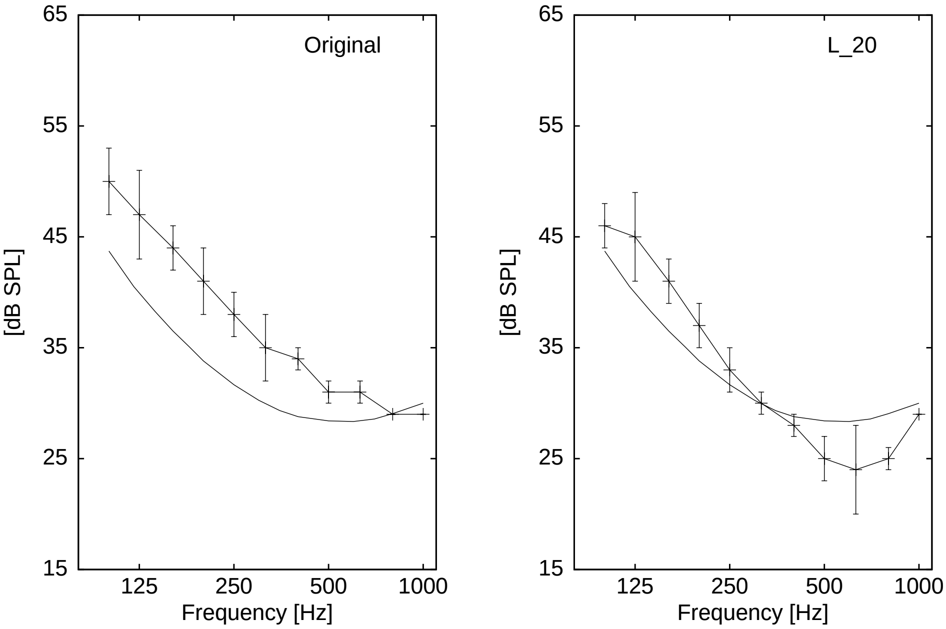

Figure 2.4.: *Left panel: Summary of the original data from Fig. 2.3 (left panel) with standard deviation. Right panel: Average of the calculated ABS values (Fig. 2.3) with standard deviation using the values of Fig. 2.3(left panel) with Eq. 2.2. Solid line represents data from ISO 226.*

next figures. Equal-loudness level contours measured by Betke [1991] (□), Fastl et al. [1990] (×), Suzuki et al. [1989] (△) and Watanabe & Møller [1990] (○) and estimated data (filled symbols) are shown in Fig. 2.5 for 20 to 50 phon and Fig. 2.6 for 60 to 80 phon.

From the data shown in Fig. 2.5 and Fig. 2.6 it can be concluded that the model can not account for all differences between the original data. Of course some data are lowered and move closer to ISO 226. For the equal-loudness level contours at high values it may be not surprising that the model fails, because it was only developed and tested for the lower contours. Data from Watanabe & Møller were determined with a procedure which differs to the one employed by Gabriel, this might be one reason for the fail of the model.

Research group	Loudness Level	Frequency Range	Method
Betke	30, 40, 50, 60	50 - 12500	constant stimuli
Fastl	30, 50, 70	100 - 1000	constant stimuli
Suzuki	20, 40, 70	63 - 12500	constant stimuli
Watanabe	20, 40, 60, 80	31.5 - 1000	adaptive 2-AFC

Table 2.2.: *The employed procedures, frequencies and loudness level contours determined by the indicated groups. The measured data are the basis for the calculation of the modeled contours.*

It is surprising that the strongest influence of the model is not seen at low frequencies, but is in the frequency region around 500 Hz. In this frequency region calculated data have a relatively sharp minimum which is not found in the measured data. This problems already occurs for calculated data from Betke (see Fig. 2.3 and Table 2.1). The shape of the weighting function is questionable (see Eq. 2.5 and Fig. 2.2). In addition, the correct calculation of REL can not be determined from these data.
In the next section the parameters of the model will be systematically analyzed.

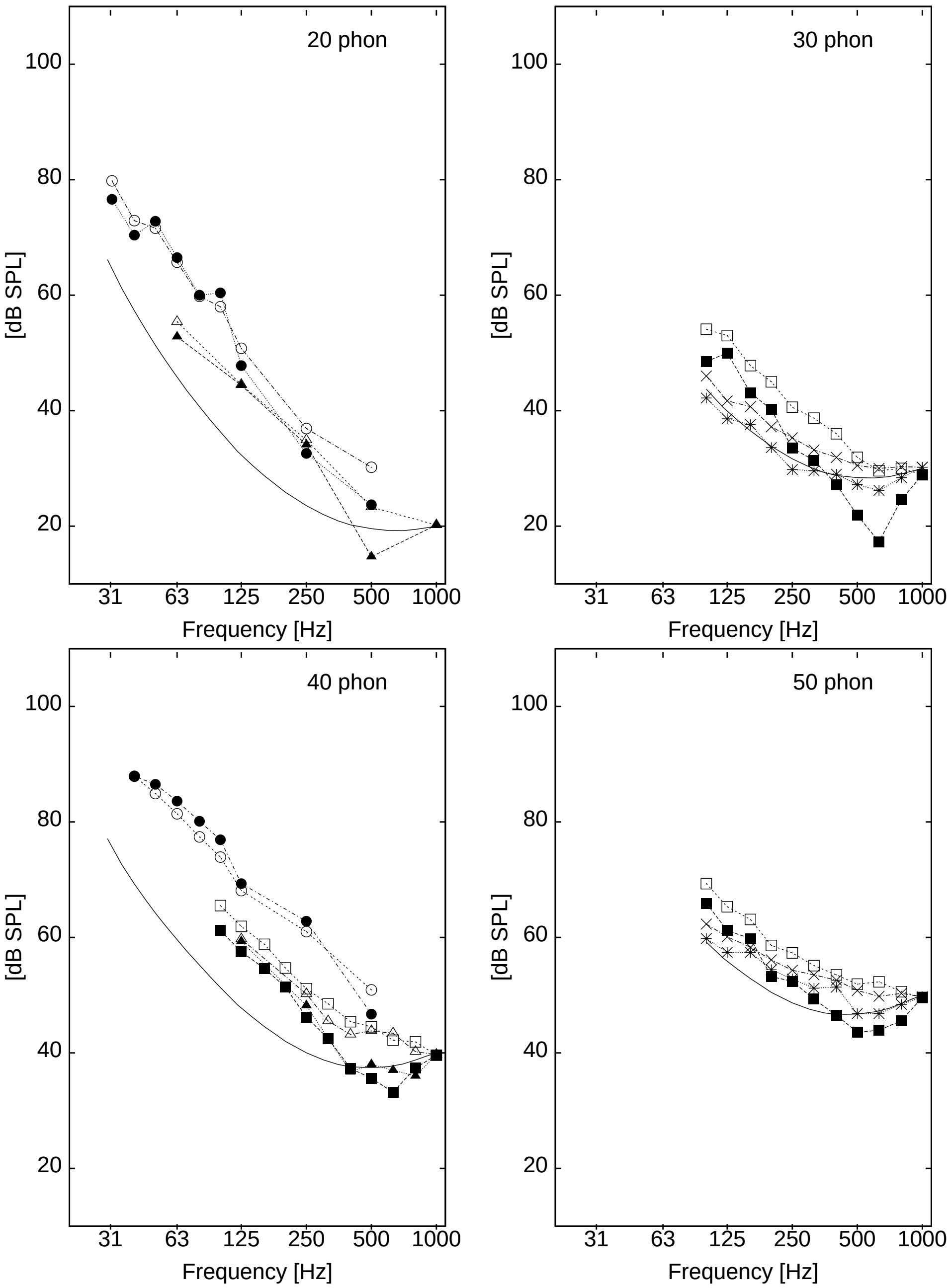

Figure 2.5.: *Equal loudness level contours measured by Betke [1991] (□), Fastl et al. [1990] (×), Suzuki et al. [1989] (△) and Watanabe & Møller [1990] (○) for 20, 30, 40 and 50 phon. Filled symbols represent ABS data calculated with the model (Eq. 2.2).*

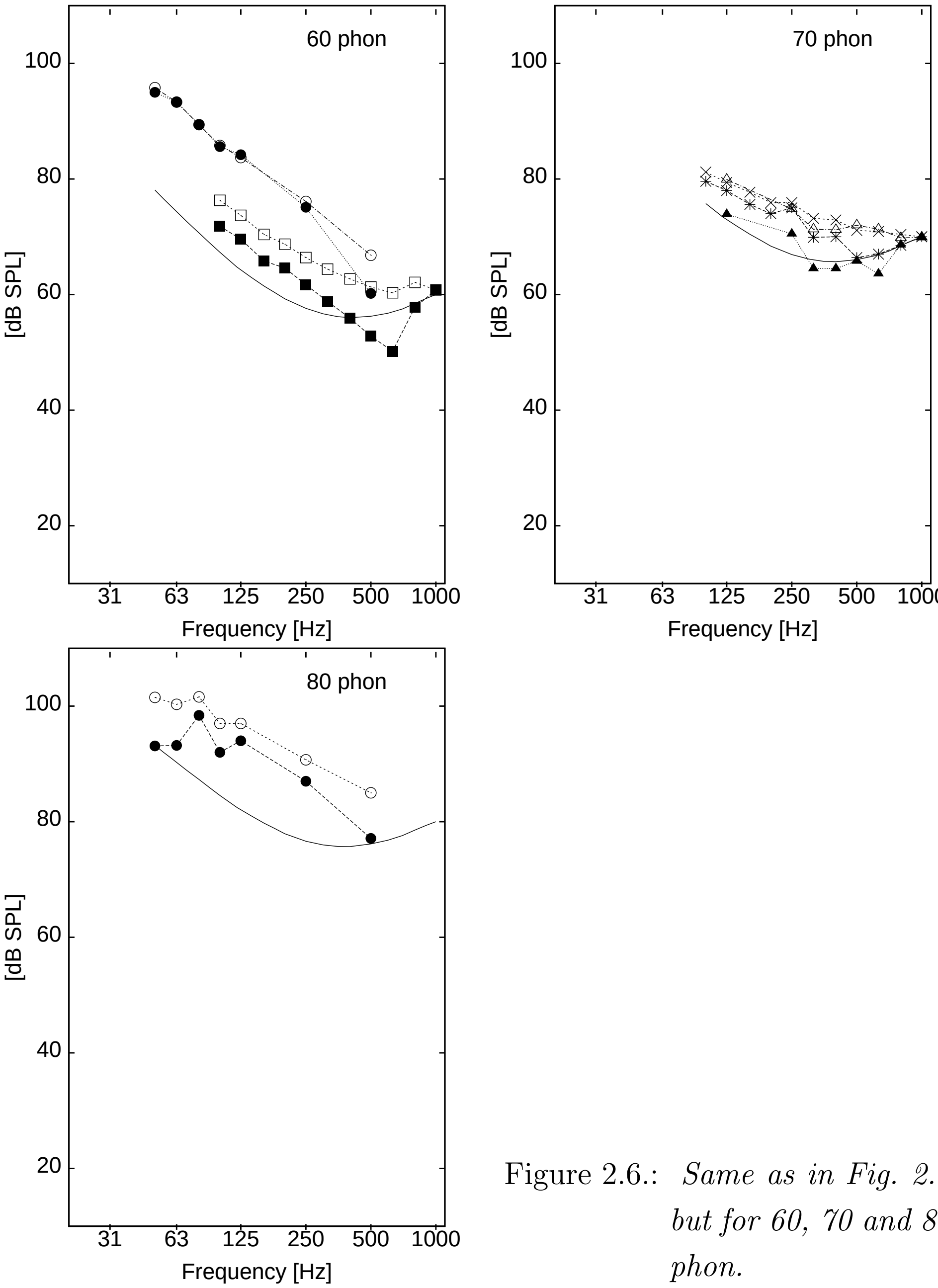

Figure 2.6.: *Same as in Fig. 2.5 but for 60, 70 and 80 phon.*

2.3. Analysis of the model

2.3.1. Influence of the determination of REL

Gabriel found that the results was mostly influenced by the highest presented level. Therefore Gabriel proposed a 80% to 20% weighting (see Eq. 2.4) for the max and min level for the calculation of REL. This weighting suggests the application of loudness levels instead of intensity in [dB SPL]. Hence, in this section the model given by the equation Eq. 2.2 was in addition applied for level values converted to Zwicker-loudness (see Zwicker and Fastl [1990]):

$$\frac{N_{1kHz}}{sone} = \frac{1}{16}\left(\frac{I_{1kHz}}{I_0}\right)^{0.3} \approx 2^{\left(\frac{L_{1kHz}-40}{dB}\right)/10} \tag{2.6}$$

The 'medium' loudness depends on the presented range and distribution of the test-signals. But the question remains what is the 'medium' loudness, relevant for the subjects' decision.

In the following REL was calculated either as medium loudness N_{50} or as a N_x, which correlates better to the loudness of time varying signals. Stemplinger & Gottschling [1997] proposed a N_5 for time varying traffic noise. A good overview can be found in Zwicker & Fastl [1999].

It must be emphasized however there are still not enough data, how the distribution of the test-signals influences the medium loudness in the determination of $ELLC$.

The optimization was accomplished with a differently chosen REL-values. Fig. 2.7 shows calculated equal-loudness level contours for different calculation of REL. The top left panel shows contours obtained whereby a REL is calculated from loudness levels is employed. All levels presented to the subject were transformed to loudness levels (Eq. 2.6) and than calculated with Eq. 2.2. The right panel shows equal-loudness level contours calculated with a N_{35} percentile as REL. Again for some frequencies (e.g. 630 and 800 Hz) the differences between the data from the model become larger than deviations between original data and for some frequencies e.g. 315 and 400 Hz, variance could be reduced. In comparison to the data shown in Fig. 2.4, the calculated values are higher than the data calculated with the original implementation of REL and lie closer to the original data. The unreasonable low data shown in Fig. 2.3 (right panel) and Fig. 2.4 (right panel) for the frequency of 630 Hz can be avoided with the employed loudness calculation. This is reflected in a smaller standard deviation at that frequency when averaging over the 4 test-series.

The figure on the bottom left side shows calculated REL-data using Eq. 2.4 but with a 50% weighting for min and max level. This calculation was done to show the difference between using the loudness (N_{50}) calculation and averaging level values (L_{50}). As one can see data calculated with the L_{50} are higher and have larger deviations than the one calculated with N_{50}. Calculated data are up to 7 dB higher than the original data (Betke 125 Hz), even for the measured data from Gabriel the calculated data are up to 4.8 dB higher than the measured data (only averaged data are shown).

Although in this section was shown that the modified calculation of REL using loudness values reduces differences between data from Fastl, Betke and Gabriel, an application to a wider range of data published in the literature was not investigated, since in the literature only little information about the presented level are available, a calculation regarding the distribution of the employed level was not possible.

In addition the question remains, if the reduction between data from the 4 studies using the proposed model leads to the correct values.

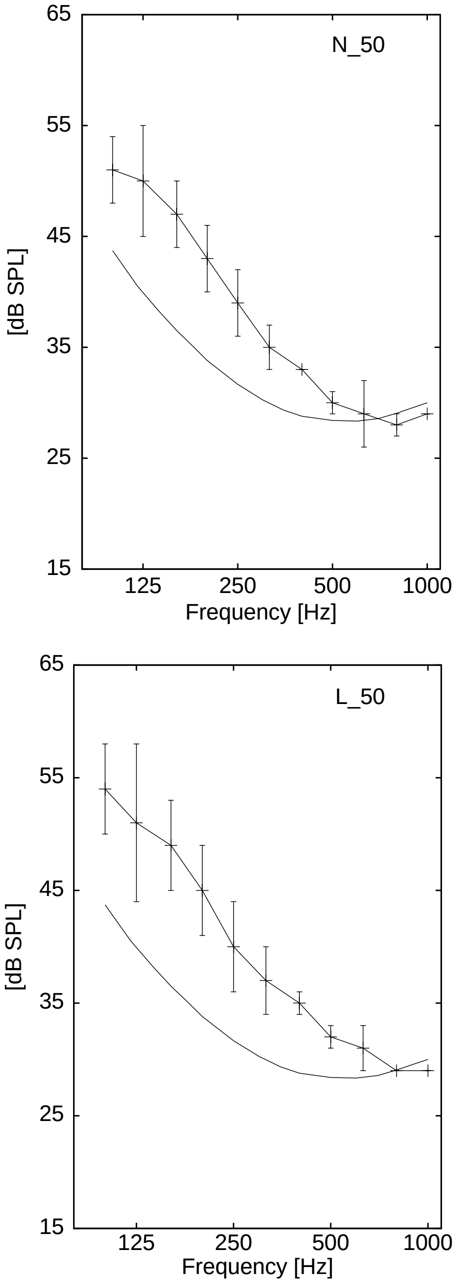

Figure 2.7.: *Top: Left panel: Average of the data calculated with loudness values (N_{50}). Right panel: Average data with standard deviation using the N_{35} percentile for the loudness calculation. Bottom: Average data with standard deviation, REL is calculated as 50% of max and min level.*

2.3.2. Influence of the weighting factor w

From Eq. 2.2 one can see, that besides REL the parameter w is necessary for the calculation of ABS. While w was just fitted for some frequencies a variation of the shape of w is not unreasonable. Eq. 2.2 can also be used for estimating the weighting factor w, when some assumptions are made for ABS.

$$w = \frac{PSE - ABS}{REL - ABS} \tag{2.7}$$

In Eq. 2.7 the problem mentioned in the previous sections is present again. No sensible statements about w are possible, when REL and ABS lie closely together. This is obviously the case when the measurement is not extremely biased. On the other hand if the measured PSE is close to the calculated ABS the interpretation of w becomes difficult again. One can see problems with the definition of the formula.

For given data for PSE and REL, changes in the ABS-value for different weighting values w can be estimated from Eq. 2.7. Taking into account the assumptions made over the existence-range of w (*i.e.* $0 < w < 1$), weighting-factors can be calculated for several frequencies. The weighting function should be relatively smooth as shown in Fig. 2.2, having one minimum at the reference frequency, with the tendency to higher weighting values at frequencies far away from the reference frequency. The range of calculated ABS values should be restricted to the presented test-tone level range. Depending on the frequency difference between test-and reference-tone the range effect should be restricted with the largest values for larger frequency differences. For a certain w, calculated ABS from the different laboratories should move closer together. Analysis was done for data from Betke [1991], Fastl [1990] and the reproduced data from Gabriel [1996] for the 30 phon curve, because the model was proposed for these data.

Results for the four different test-series, w and dependent ABS-values are shown in Fig. 2.8 to Fig. 2.10 for test-tone frequencies in the range from 100 to 1000 Hz. REL was calculated from Eq. 2.4 as proposed by Gabriel.

As one can see from these results it is still not clear in which way the different parameters influencing the ABS-value should be weighted to minimize the variance of the data. For example in Fig. 2.9 (400 Hz), a large weighting-factor causes a breaking apart of the different test-series. A similar effect is found for a factor which is too small. In the present case a value of w=0.5 would be the optimum. In this way an optimization can be accomplished to the weighting-factors, which corresponds for this frequency well to the values in Gabriel [1996] (e.g. w= 0.5 for 400 Hz calculated using Eq. 2.5). However, for some frequencies, e.g. 100 Hz (Fig. 2.8), it is not possible to determine an optimum weighting-factor. Because of the experimental contexts no better agreement of the contours could be achieved.

It can be seen from Eq. 2.3 that the calculation of ABS obviously leads to problems: particularly for large distances between test- and reference tone ($w \rightarrow 1$), the denominator becomes small. Fundamentally, the PSE is determined from the subjects' answers and has an uncertainty which is reflected by the different answers when repeating the experiment. The intraindividual measurement error is a measure for the reproducibility of the same experiment several times for one subject.
When averaging the results across different subjects the interindividual standard deviation gives a measure how well a group of subjects agree in their answers. These 'errors' always enlarge ABS (Eq. 2.3) because of $0 < w < 1$. The weighting-factor w should always be positive. At large weighting factors ($w \rightarrow 1$) these unavoidable 'measurement errors' increase dramatically. This means for large frequency distances between test- and reference-tone where the task for subjects is most difficult, the model enlarges deviation in an unreasonable way.
It is not clear if this simple model could account for all parameters affecting the range, since variance already occur between original data and reproduced measurements (see Fig. 2.3, left panel). Hence additional factors to the range-effect might influence the subjects answers.

A difference in Gabriel's experiments from the original experiments of Fastl is for example the distribution of the presented test-tones. Whereas Gabriel presented uniform distributed level, in the original work of Fastl levels ± 1.9 dB apart from a starting level are presented 20 times, ± 4.9 dB 10 times and ± 7.9 dB only 5 times. However, in an experiment performed by Gabriel no influence of the distribution of the presented signals was found. Another possible explanation for deviations between results from different laboratories was given in the work from Gabriel [1996] who argued, that additionally, half of the group of listeners even judged all test-tones presented with frequencies below 400 Hz and levels chosen as by Fastl et al. to be softer than the reference tone. Therefore, other factors than the range, e.g. the group of listeners or level range size most probably have contributed to the low values measured by Fastl et al.. Because these assumptions are not systematically tested, the reason of this effect is still unclear.

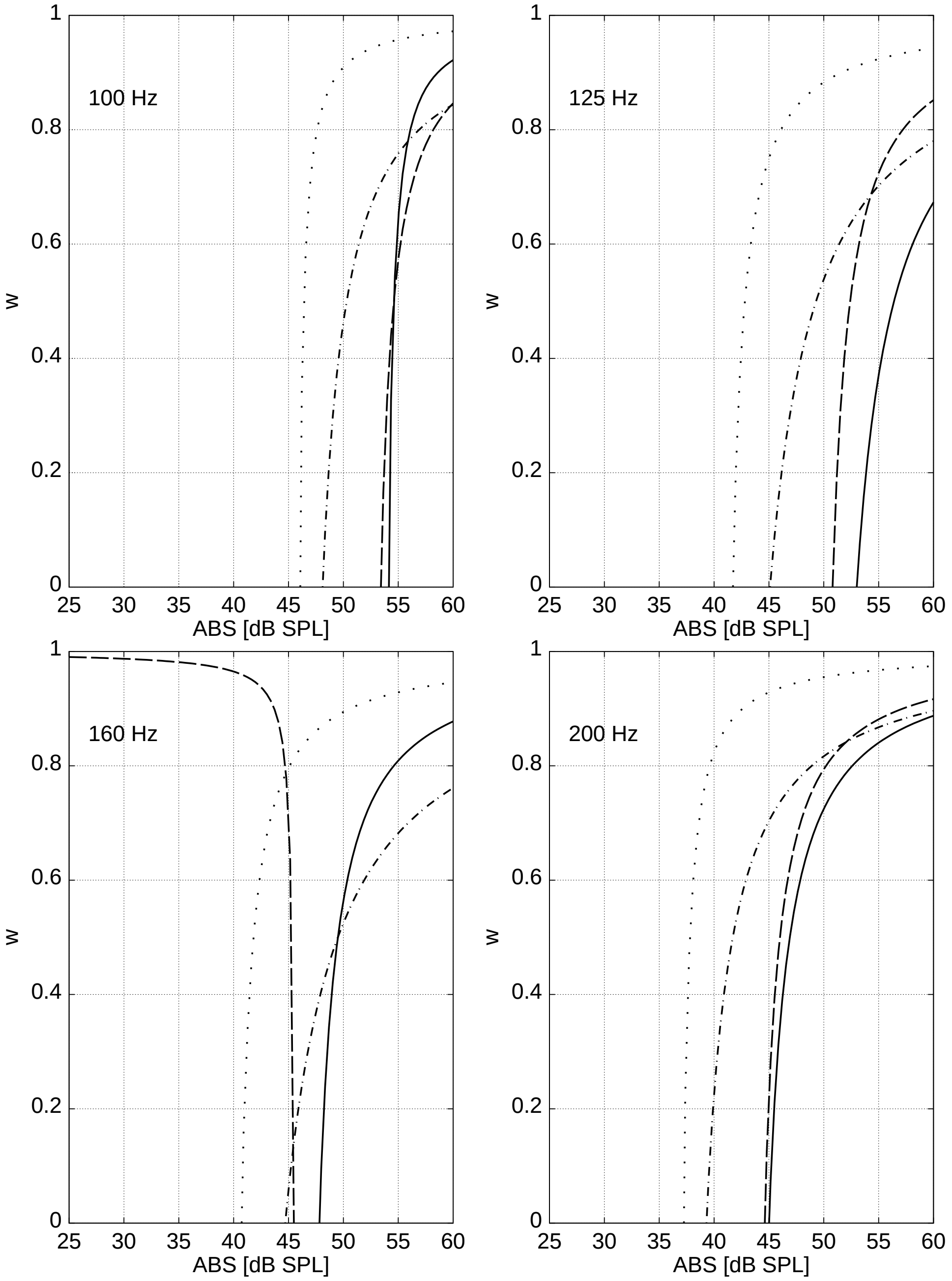

Figure 2.8.: *Dependency between weighting factor w and ABS-value for the test-tone frequencies from 100-200 Hz.(see Eq. 2.7). An existence area for w between 0 and 1 is assumed. Betke (—), Betke-BG (- - -), Fastl (· · ·) and Fastl-BG (· − ·)*

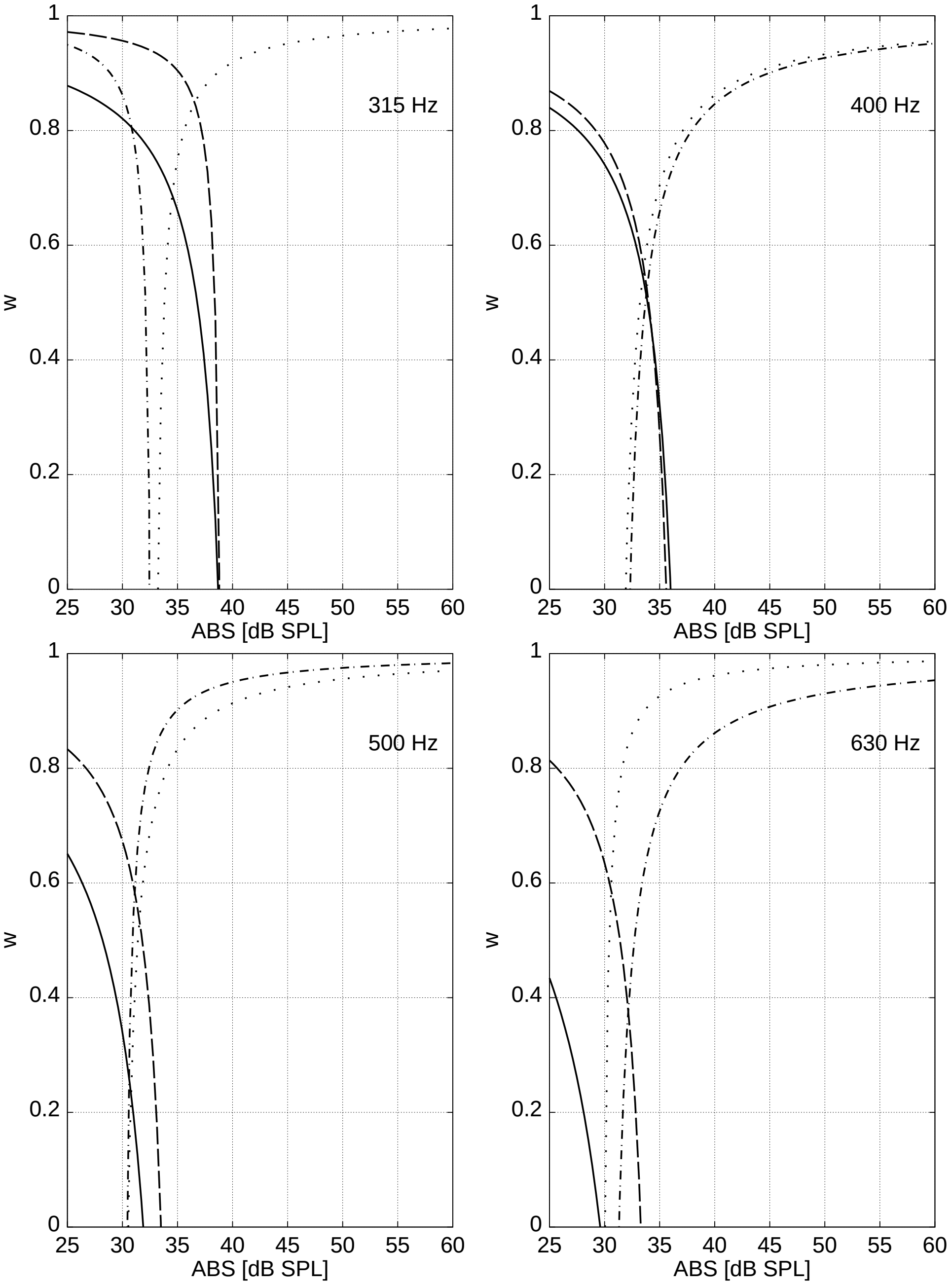

Figure 2.9.: *Same as Fig. 2.8 but test-tone frequencies from 315-630 Hz.*

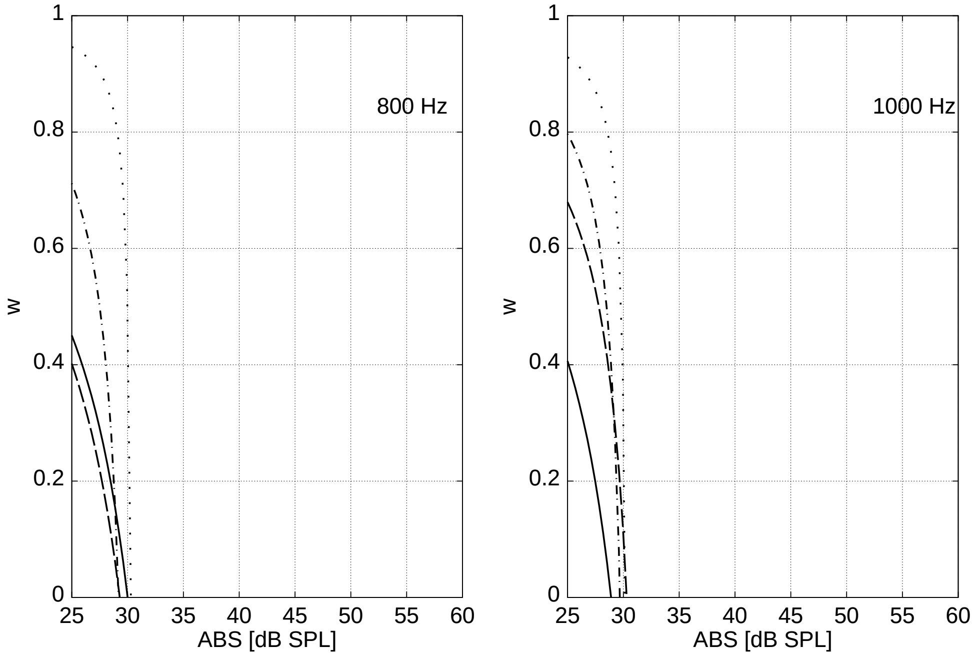

Figure 2.10.: *Same as Fig. 2.8 but test-tone frequencies from 800-1000 Hz.*

2.4. Summary

The simple model used in this study is not able to predict the range effect described by Gabriel. Applied to data from the literature, the large variance between different laboratories could only partly be minimized. For higher levels and for other procedures employed for the determination of equal loudness level contours (e.g. Watanabe [1990] used an adaptive procedure) this is not surprising because the model was fitted to the 30 phon curve measured with the method of constant stimuli. Even for these curves the model was not able to explain all differences. Although for some frequencies the tendency of the calculated $ELLC$ to move closer together could be observed the contrary effect also occurred (e.g. 630 Hz). Unfortunately, systematic data on the frequency dependent weighting w are still missing. The calculation procedure of the relative value REL, which depends essentially on the experimental design is not yet known. Only further experimental data will be able to contribute to the validation of Eq. 2.2. A substitution of the proposed calculation of REL by using loudness values and a loudness calculation using the N_{35} percentile could only avoid unreasonable low data around 630 Hz. For some frequencies (e.g. 630 and 800 Hz) the differences between the data from the model were larger than deviation between original data. On the other hand variance can be reduced for frequencies of 315 and 400 Hz. In comparison to the data calculated with the original model, these values are at higher levels and lie closer to the measured data.

Discrepancies between the original Betke [1991] and Fastl et al. [1990] data and the reproduced data from Gabriel [1996] which were obtained under equivalent measurement setup could not be explained by the model. Though at some frequencies they are almost as high as the effect caused by the different test-tone level ranges. One reason might be that different groups of subjects which where not the same in the original and the reproduced measurements produce different results and that the amount of subjects was relatively small.

Summarized, the original model as well as the modified model was not able to explain difference between data from different laboratories. In the next chapter a different measurement setup will be employed for the determination of equal-loudness level contours taken into consideration the results of Gabriel to minimize range effects.

3. Comparison of the method of constant stimulus with an adaptive procedure

When using the method of constant stimuli for the determination of $ELLC$ the point of subjective equality (PSE) is influenced by the experimental setup. Especially, the absolute position of the employed test-tone level range affects the results. Modeling the range effects by a simple adaptation model could only explain the measurement bias to a certain amount (cf. previous chapter). A better way than modeling measurement errors is to avoid them by using a modified measurement procedure, which takes the findings of Gabriel's [1996] work into account. Since it is impossible to totally avoid measurement errors the aim is to minimize and to quantify the influence of the measurement setup on the results.

Based on the results of Gabriel [1996] some assumptions concerning the measurement setup can be made to minimize range effects. Because the most critical parameter for measurements of equal-loudness levels contours using the constant stimulus procedure are the absolute position of the employed test-tone level range and high test-tone levels (see Eq. 2.4), these parameters should be avoided. A possible way to reduce the influence of these critical parameters is to use an *adaptive* two-alternative forced choice procedure (2-AFC), with a starting point close to the expected PSE. An adaptive 2-AFC procedure procedure is in accordance with ISO/ TC43/

WG1/N122 [1988] for the determination of $ELLC$. Adaptive in this context means that the test-tone starts at a certain level and the level is increased or decreased in the next presentation depending on the answer of the subject. The amount of the in- or decrease is reduced after a distinct number of reversals of the level sequence. After the step-size reached a minimum size, it is held constant for an additional amount of reversals. This procedure reduces the number of high levels. In addition, most of the presented levels are close to the PSE. The PSE can be obtained out of a set of trials at the end of the experiment where the test-tone levels are near the PSE and temporal distance for the presentation of high levels is large. The adaptive procedure was already used in other studies for the determination of equal-loudness level contours, (e.g. Watanabe [1990], Lydolf and Møller [1997] and Lydolf [1999]).

Since a systematic investigation of the influence of the experimental design on the results was not done in these works, in this chapter several experiments using an adaptive 2-AFC procedure are presented. In addition, a comparison between the employed methods is missing in the literature. Especially, it is of interest if something similar to the range-effect for the method of constant stimuli exists for an adaptive 2 AFC procedure. In the first experiment (section 3.2.1) it is investigated if the both test procedures, the method of constant stimuli and the adaptive 2-AFC procedure, lead to systematic different results when the same signals are presented in the same order. The experiment described in section 3.2.3 investigates how the measured PSE is changed when the order of the presented test-tone is altered. In section 3.2.4 the influence of a small level displacement of the presented test-tone levels using the method of constant stimuli is quantified. This experiment is motivated by an experiment described in the work of Betke [1991] where it was shown that a small (2.5 dB) change of the presented test-tone level range could already effect the results. Especially for the lower equal-loudness level contours and in the frequency range between 250 and 630 Hz the results are shifted up to 2.8 dB. The shift was originally employed between two level ranges to increase the level resolution. Therefore two level ranges were presented from which one was shifted by 2.5 dB. In the experiments from Betke the method of constant stimuli with a step-size of 5 dB was employed. In the last experiment in this chapter the influence of free parameters in the adaptive procedure is systematically varied to quantify the amount of the influence of the experimental design on the resulting PSE and to compare the procedures.

3.1. Measurement setup

3.1.1. Apparatus

Signals are generated with a Sun Workstation ELC and digital-analog (D/A -16 bit) converted by a PC with an Ariel DSP 32 C board. Signals are amplified (Pioneer A.602-R) and presented via headphones (Stax SR λ). The experiments are performed in a sound-attenuating booth. All devices except the PC-keyboard for the answers of the subjects are placed outside to avoid background noise. The computer display is placed in front off a window of the sound-attenuating booth. The experimental setup is schematically shown in Fig. 3.1. During the measurements the following text is displayed on the screen (see Appendix A.2 for the original text in German):

> Which of the both signals is louder?
>
> Press key $\boxed{1}$ if the first signal is louder.
>
> Press key $\boxed{2}$ if the second signal is louder.

After the subject had given its response via PC-keyboard the next pair of signals was presented.

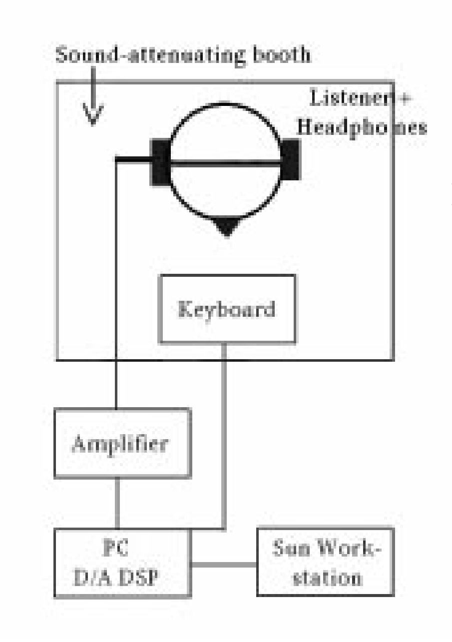

Figure 3.1.: *Schematic representation of the employed measurement equipment for the determination of equal-loudness level contours.*

The headphones' transferfunction is measured using an artificial head (Head Acoustics HMS II.2) and is shown in Fig. 3.2. In the range from 200 to 1000 Hz (which is examined in this work) the curve is relatively smooth. The small deviation from a desired sound pressure level is equalized in the employed software. The headphones are seated twice on the artificial head to examine the reproducibility and the measurement is done for both sides individually. In the frequency region from 200 to 1000 Hz the average difference between the two measurements is $\leq$ 0.7 dB. The difference between right and left ear is $\leq$ 0.9 dB.

3.1.2. Adaptive procedure

The sinusoidal test-tone frequencies are 200, 400, and 630 Hz. Test- and reference-tone are presented in random order. The reference tone (frequency 1000 Hz) has a level of 30 dB. Test- and reference tone have a duration of 1 s and are separated by a silent interval of 500 ms between the signals. Signals are gated with 50 ms Hanning ramps on and off. A

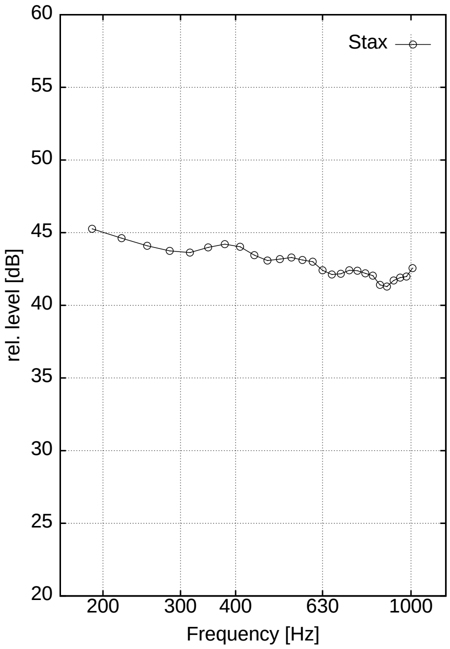

Figure 3.2.: *Transferfunction of the Stax headphone in the frequency range from 200 to 1000 Hz. The ordinate represents level as a function of frequency.*

written instruction is given to the subjects (see Appendix A.3). At the beginning of this experiment they perform at least 3 training runs with the employed frequencies. Subject's task is to decide which of the two presented tones is louder and to give the response on a keyboard by pressing 1 or 2, if the first or second tone respectively, is louder.

In the adaptive procedure the starting level of the test-signal at the beginning of an experimental run varied within the different experiments. The initial step-size is 8 dB. After the first reversals in the track, the step-size is halved. Minimum step-size is 2 dB and after 4 reversals with minimum step size the measurements stops (Fig. 3.3). As the point of subjective equality the median over these four last reversals is calculated.

Instead of taking more reversals when measuring at the smallest step-size, the measurement is repeated four times. There are two reasons for doing this. First, when determining the point of subjective quality, the subject guesses at the 50 % point which of the tones is louder and a longer period of doing this is not motivating. Second, when the track ends in a local mini- or maximum which might be due to a lack of attention, the influence of such a result is reduced.

For the method of constant stimuli the level range is chosen by the investigator. A detailed description is given in 'Results' precedes each experiment.

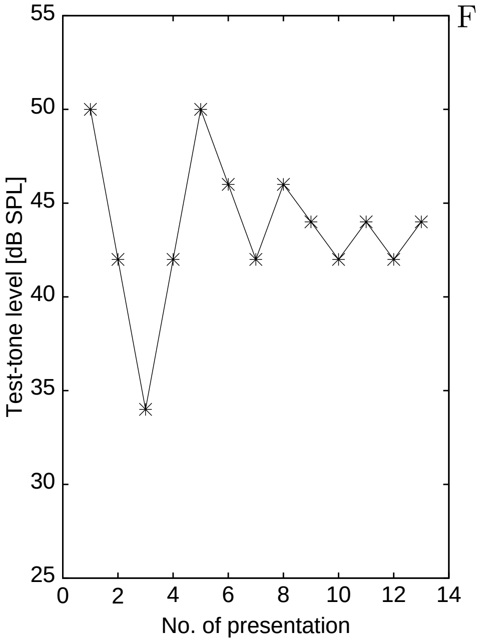

Figure 3.3.: *Schematic presentation of one track for a test-tone in an adaptive 2-AFC procedure. Starting level for the test-tone is 50 dB. Initial stepsize is 8 dB. The reference signal which has a constant level is not shown here. It is presented in random order before or after the test-tone.*

3.1.3. Listeners

Six subjects, three male and three female, participate in the experiment. They are all students and experienced in psychoacoustic experiments. The age of the listeners ranges from 24 to 30. All subjects are tested with an pure tone Békésy audiometer Brüel & Kjær 1800. Test-tone frequencies are 500, 1k, 2k, 3k, 4k, 6k and 8 kHz as recommended in ISO/TC43/WG1/N122 [1988]. All subjects show a hearing loss less than 10 dB for frequencies below 4 kHz.

3.2. Results

3.2.1. Comparison of measurement procedures under comparable conditions

One of the main differences between an adaptive AFC procedure and the method of constant stimuli is the influence of the investigator and the subject on the temporal order of presentation and the level range of test-tone levels in the course of the experiment. Whereas in the method of constant stimuli the order and the level range is completely set by the investigator, in the adaptive procedure this influence is strongly reduced. In the adaptive procedure the investigator must set in advance starting level and step-size, whereas subjects' answer determines the levels of the test-tones. In order to investigate whether this difference affects the PSE, first an adaptive AFC measurement is performed.

3.2.2. Adaptive procedure versus constant stimuli

Individual equal-loudness levels for three frequencies are measured for six subjects with the adaptive 2 AFC procedure. Starting level in the adaptive procedure are 55 dB (200 Hz), 45 dB (400 Hz) and 40 dB (630 Hz). These values are 10 dB higher than the PSE obtained by Betke [1991] for the 30 phon curve. Before the final measurement at least 3 training sessions are done. The individual data, i.e. order and level of the presented test-tones for each subject are recorded for the next experiments. In order to test the influence of the employed measurement procedure, (adaptive versus constant stimuli), the individual test-tone levels and order obtained from the adaptive experiment are presented to the subject. In contrast to the adaptive AFC experiment, this time the method of constant stimuli is employed. The recorded levels obtained in the adaptive experiment are presented again, but in this experiment the level of the next presented test-tone is independent of the subject's answer. Each subject repeated this experiment 4 times. Subjects were not informed about the different procedures.

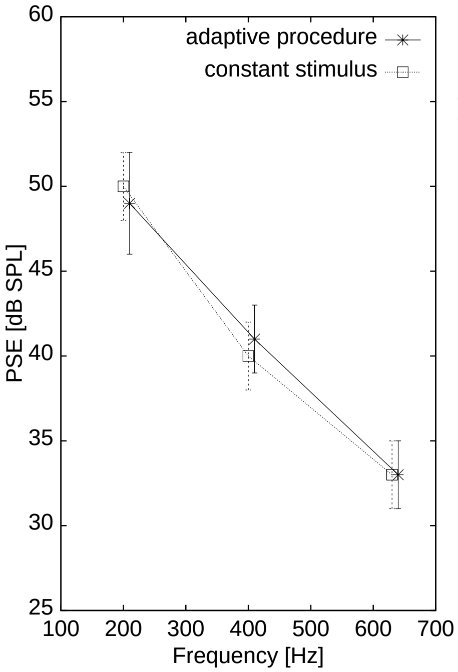

Figure 3.4.: *Mean PSE averaged across listeners of the results obtained with the adaptive procedure and the method of constant stimuli. Mean PSE and standard deviation for three frequencies are shown.*

In Fig. 3.4 mean PSE and standard deviation obtained with the adaptive procedure ($*$) and with the method of constant stimuli ($\square$) are plotted. The curve shows the typical shape of an equal-loudness level contour, going from high levels at low frequencies to lower level at medium (630 Hz) frequencies reflecting the sensitivity of the human ear. For both procedures, interindividual errors are by a factor of two larger than intraindividual errors (not shown).

Reproducibility of the experiments is almost independent of the presented test-tone frequency. There is a tendency of smaller interindividual errors of about 1 dB at frequencies closer to the reference.

The differences in the PSE values obtained with the two procedures are smaller than 0.9 dB, i.e. much smaller than the interindividual difference. The difference between the methods is not statistically significant

(Wilcoxon Test, $p > 0.05$) for any of the frequencies. Thus, no influence of the measurement procedure can be observed. This finding is consistent with the impression of the subjects. None of the subject reported any differences between the procedures in a short interview after the experiment. This is probably because subjects performed the method of constant stimuli with their individual test-tone levels taken from the adaptive measurement. It is remarkable that differences in PSE between the frequencies are of the same amount as the ones measured by Betke [1991]. Although a direct comparison to the data from Betke which were obtained under free-field conditions is not meaningful .

3.2.3. Influence of the temporal order of test-tone level in constant stimulus experiments

This experiment is performed in order to test if the temporal order of the presented test-tones influences the results when using the method of constant stimuli. Again the individual levels obtained in the adaptive experiment are used for each subject. An ascending order of the presented test-tone levels is used, i.e. test-tone levels are sorted starting from low levels to high levels. In comparison to the adaptive procedure, levels far apart from the resulting PSE are presented more at the end (cf. Fig. 3.5 (left panel)).

Mean PSE and standard deviation of ordered test-tone level sequence from experiment 3.2.3 indicated by a ($\triangle$) and for comparison the results from the adaptive procedure (experiment 3.2.2 indicated by a ($*$)) are shown in Fig. 3.5. In comparison to the adaptive procedure employed in experiment 3.2.2 the resulting PSE are up to 3.2 dB lower. This might be interpreted as a temporal effect of the presented test-tone levels. The difference between the procedures is only for the frequency of 400 Hz statistically significant (Wilcoxon Test, $p < 0.05$). Another observation is that the interindividual standard-deviations increase by a factor 1.5 for the ordered test-tone level sequence in comparison to the adaptive proce-

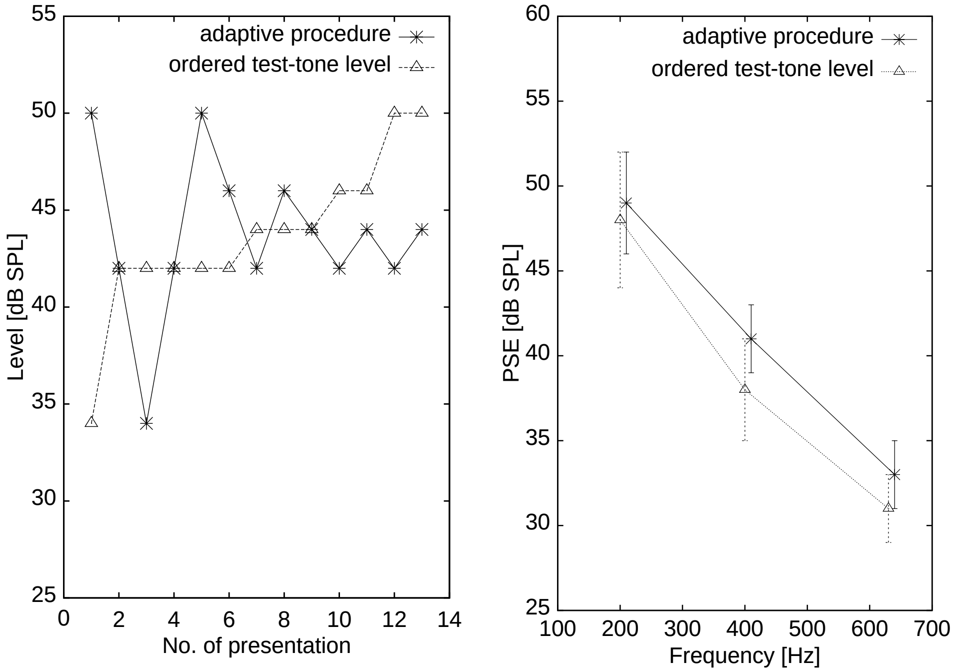

Figure 3.5.: *Left: Test-tone level from Exp. 3.2.2 and ordered test-tone levels starting from low going to high levels for one subject. Right: Mean PSE and standard deviations for six listeners of Exp. 3.2.2 and Exp. 3.2.3 for three test-tone frequencies.*

dure. The intraindividual standard deviations are in average smaller than 2 dB.

3.2.4. Influence of the position of the level range in constant stimuli experiments

The underlying idea goes back to experiments performed by Betke. The method of constant stimuli with a step size of 5 dB is used in these experiments. Two level ranges are shifted by 2.5 dB to increase the level resolution. Betke observed for the 40 phon contour in the frequency range between 250 and 630 Hz that small deviations of the presented test-tone level range (in Betke [1991] 2.5 dB) lead to a displacement of same amount

in the resulting PSE. Large deviations however, contribute less.
In contrast to the experiment in the present study, Betke presented the same equally distributed test-tone levels for all subjects. In the present study the individual order of test-tones obtained from Exp 3.2.2 is used for each subject. The measurement procedure is the method of constant stimuli. To test if this effect also occurs with the individual data obtained from adaptive procedure (Exp. 3.2.2), these test-tone levels are shifted by +2 dB and - 2 dB respectively.

The results of Exp. 3.2.4 are plotted in Fig. 3.6. The difference between the results is around 2 dB and has a tendency to decrease for test-tone frequencies lying closer to the reference-tone. In contrast to the results from Betke which are obtained with equally distributed test-tone levels, the shift of 4 dB of the test-tone levels only shifts the resulting PSE by about 2 dB. The difference between the shifted ranges is only for the frequencies of 200 and 400 Hz statistically significant (Wilcoxon Test, $p < 0.05$). The smaller influence in comparison to the data from Betke could be a consequence of the individual test-tone levels for each subject used

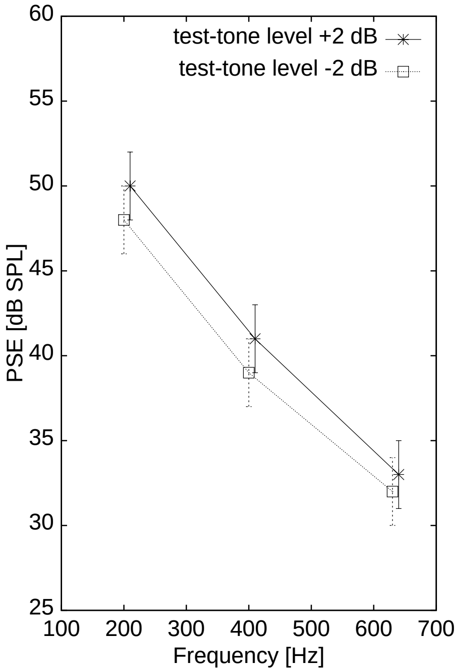

Figure 3.6.: *Mean PSE and standard deviations for six listeners of Exp. 3.2.4 for three test-tone frequencies (slightly shifted in the figure). Presented test-tone level are shifted +2 dB ($*$) respectively by -2 dB ($\Box$).*

in the present study. Experiments with larger shifts of the test-tone levels were not performed, because some subjects already claimed that the test-tone is always louder than the reference-tone, or vice versa, depending on the applied level shift.

3.2.5. Influence of the starting level in the adaptive 2-AFC procedure

In this experiment the influence of the starting point on the resulting PSE is determined using an adaptive 2-AFC procedure. Besides the initial step-size this is the main parameter which has to be set by the investigator prior to the measurement. From the experiments performed by Gabriel it is known that test-tone level far above the resulting PSE have the strongest influence on the result (see Eq. 2.4). An increase of the resulting PSE can be assumed when increasing the starting point of the test-tone level. The amount of the effect can not be predicted by the model, described in section 2.1, since it is obtained with a different procedure, i.e. the method of constant stimuli. The influence of the starting level in an adaptive procedure can be compared with the influence of the absolute position of the employed test-tone level range in the method of constant stimuli. While the amount of the influence is known for the method of constant stimuli from experiments from Gabriel, in this chapter it is quantified for the adaptive procedure.

The starting levels for the test-frequencies 200, 400 and 630 Hz are 20, 40 and 70 dB. All other parameters such as initial step-size of the procedure are the same as in Exp. 3.2.2. The PSE value is averaged across the 6 subjects who repeated the experiments 4 times.

The results for three different starting levels and three different test-tone frequencies are shown in Table 3.1 and Fig. 3.7. Different symbols indicate PSE for the employed test-tone frequencies: 200 Hz ($\triangle$), 400 Hz ($\times$) and 630 Hz ($\square$). In agreement with the findings for the method of constant stimuli, PSE increases with increasing starting level. This is observed for

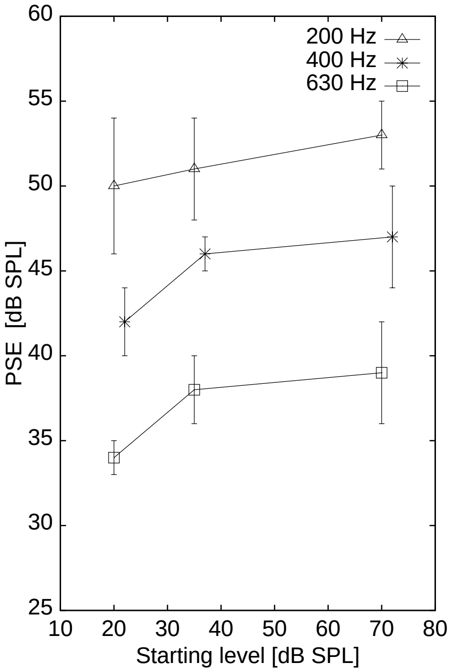

Figure 3.7.: *Mean PSE and standard deviation for 3 test-tone frequencies. The abscissa indicates the starting levels for the adaptive procedure.*

all test-tone frequencies.

The largest influence of the starting level is about 5 dB at 630 Hz, the smallest effect is at 200 Hz with 2.6 dB. The differences between the starting level of 20 and 40 dB are larger than the differences between 40 and 70 dB. The influence of the starting level is statistical significant for of 400 and 630 Hz (Friedman Test, $p < 0.05$). For the test-tone frequencies of 400 and 630 Hz, differences from 20 to 40 dB are statistical significant (Wilcoxon Test, $p < 0.05$), whereas a change from 40 to 70 dB is not significant (Wilcoxon Test, $p < 0.05$).

It is remarkable that the interindividual errors are larger than intraindividual errors (not shown) by a factor of about two. This means that every individual subject can repeat the measurements with a very good reproducibility, but the PSE of the six subjects is different. Inter- and intraindividual errors are almost independent of the distance from the reference. A slight increase of the intraindividual error of about 1 dB is observed for frequencies far away from the reference.

In comparison to the method of constant stimuli, results obtained with the

f [Hz]	20 - 40 dB	40 - 70 dB
200	0.8	1.8
400	4.3	0.5
630	4.5	0.5

Table 3.1.: *Difference between starting level 20 and 40 dB and 40 and 70 dB for 3 test-tone frequencies. Initial step-size 8 dB. Level of the reference is 30 dB.*

adaptive procedure are less influenced by the experimental setup, as long as an appropriate starting level is chosen. In the adaptive procedure a shift of 50 dB of the starting level shifts the resulting PSE by about 5 dB, whereas a shift of 30 dB of the presented level range shifts the results by 13 dB when using the method of constant stimuli (Gabriel [1996]).
A further advantage of the adaptive procedure is that levels far apart from the resulting PSE are always presented first, which means that they are presented long before the adaptive procedure converges to the point of subjective equality.

3.3. Discussion

Equal-loudness level contours obtained with the method of constant stimuli are influenced by the choice of the presented test-tone level range (see Gabriel [1996]). In the previous chapter was shown that the model could only explain some differences between data from the literature. In this chapter a measurement procedure is employed which takes the results from Gabriel obtained with the method of constant stimuli into account. The aim is to propose a measurement procedure which avoids the influence of the measurement setup on the result. While this is probably not possibly in a perfect way, the influence of the procedure should be minimized and quantified. From the results of Gabriel it is known that the results are most influenced by the position of the employed test-tone level range and the main shift in the results is caused by high test-tone level (see Eq. 2.2 and Eq. 2.4).

The choice of the employed measurement procedure is restricted to a two **A**lternative **F**orced **C**hoice (2-AFC) procedure, recommended by ISO/ TC43/ WG1/N122 [1988] an adaptive 2-AFC procedure is employed to reduce the influence of the procedure on the resulting PSE.

In section 3.2.2 the influence of the employed procedure is tested. The employed procedure, adaptive vs. constant stimuli, has only a small effect on the resulting PSE when using individually obtained test-tone levels.

The PSE measured with the method of constant stimuli depends not only on the absolute position and the presented max and min test-tone levels but also on the order of the presented test-tone levels (Exp. 3.2.3). When presenting test-tone levels, in a constant stimuli experiment, in the way that high level are presented at the end, the resulting PSE is up to 3.2 dB lower, although test-tones are individually obtained with an adaptive procedure.

Results from Betke [1991] show that a small shift in the presented test-tone level range by 2.5 dB in the method of constant stimuli shifts the resulting PSE by the same amount. As observed by Betke a small shift of the presented test-tone level, leads to a displacement of the PSE (Exp. 3.2.4). However, the effect is smaller than in the data from Betke. The results of Exp. 3.2.4 show only a shift of 2 dB of the PSE values, for a shift of 4 dB of the presented level. In comparison to Betke, individual data obtained from an adaptive measurement are presented.

In Exp. 3.2.5 the influence of the employed starting level is measured for the adaptive procedure. For the adaptive procedure it is shown that depending on the presented starting level the PSE values are shifted upwards, with increasing starting level. The largest influence of the starting level is 5 dB for a level difference of 50 dB at 630 Hz. The starting level corresponds to the shift of the level range in the method of constant stimuli.In the experiments of Gabriel [1996] using the method of constant stimuli a shift of 30 dB of the presented level range leads to a shift of 13 dB for frequencies far away from the reference. This indicates that the results obtained with an adaptive procedure are less influenced by the experimental setup. In contrast to the results which are obtained using the method of constant stimuli (from Gabriel [1996]) the effect is almost independent of the frequency with the slight tendency of getting smaller for frequencies far away from the reference. In the results from Gabriel the bias is of the same amount from 125 to around 630 Hz and than decreases, with a minimum around the reference frequency.

Another reason for the almost frequency independent influence of the starting level might be the following aspect. The PSE for 200 Hz is around 15 dB higher than the one for 630 Hz, this means even for a starting level of 70 dB after 3 steps down with a step-size of 8 dB the level gives a less loud impression whereas for 630 Hz 5 steps of 8 dB are needed. The indication of the difference between the starting level and the level of the reference in [dB] might be not adequate. A better description might be based on loudness level or on in comparison to the measured PSE.

Using an adaptive procedure the starting point should be selected carefully because it still influences the resulting PSE.

One problem when determine the influence of the starting level on the resulting PSE in this experiment for the 30 phon contours is that the presented test-tone level are close to the threshold of hearing. Therefore in the next chapter a level of 50 dB is used for the reference. Another advantage is that the level of the reference is in the middle and the dynamic range for the test-tone level and can be placed more symmetrically around the reference level.

3.4. Conclusion

Measuring $ELLC$ with the method of constant stimuli and an adaptive 2-AFC procedure lead to the same result when using individually obtained test-tone levels presented in the same order. However, the results using the method of constant stimuli are still influenced by the order of the test-tone level and by a small displacement of the level range even when the individually obtained test-tone levels are employed. Using an adaptive procedure the influence of the starting level is much smaller than the influence of the level range in the method of constant stimuli. However, the results in this chapter are only obtained at a relatively low level, therefore the variation of the starting level is limited. In the next chapter this limitation will be avoided by using a higher reference level.

4. Influence of the measurement setup on the result for the adaptive procedure

4.1. Introduction

In the previous chapter equal-loudness level contours obtained with an adaptive 2 alternative forced choice procedure (2-AFC) are compared with the method of constant stimuli. It is shown that the largest influence of the starting level is 5 dB for a level difference of 50 dB at 630 Hz. In the experiments of Gabriel [1996] using the method of constant stimuli a shift of 30 dB of the presented level range leads to a shift of 13 dB for frequencies far away from the reference.

Thus, the adaptive procedure leads to results which seem to be less biased by the presented test-tone level. In the adaptive procedure several parameters must be chosen by the investigator. However, in the previous chapter the influence of only one methodical parameter (the starting level) is investigated in a restricted range. The starting level is probably the most important parameter. Apart from that parameter also the initial step-size might affect the results. In this chapter the influence of these two parameters on the results is quantified in order to propose a preferred choice of the parameters.

In the previous chapter one frequency region of the 30 phon contour is measured. A restriction of measuring the 30 phone contour is that the presented test-tone levels can get close to the subject's threshold of hearing after a few answers. Therefore in this chapter a level of 50 dB is used for the 1 kHz-reference. In addition, since the level of the reference is more in the middle of the dynamic range of the ear, the test-tone level can be placed more symmetrically around the reference level.

While differences between data from different laboratories are observed for low as well as for the higher equal-loudness level contours, it is no restriction to measure the influence of the experimental setup for the 50 phon contour. Hence, a comparison with the results of the 30 phon contour measured in the last chapter and with data from the literature can be done. Fig. 2.6 shows also large differences between data from the literature of other loudness contours, e.g. differences of the 50 phon contours are as large as of the 30 phon contour.

The influence of different starting levels is quantified for different initial step-sizes and a suggestion is made for a less biased measurement setup. The difference between intra- and interindividual differences is examined more closely. An attempt is made to explain these differences. In addition it is investigated if more central or peripheral processes in the auditory system are involved in the bias effects.

4.2. Methods

The same apparatus as described in Section 3.1.1 is used for the experiments. The PSE is measured with an adaptive 2-AFC procedure. Signals are presented diotically via headphones if not stated otherwise. As the reference a 1000 Hz tone is used. The test-frequencies are 200, 400, 630 and 1000 Hz, i.e. the highest test-frequency is equal to the reference frequency. The level of the reference is 50 dB SPL. Different initial step-sizes are used in the experiment ranging from 4 to 16 dB. As the starting level 30, 50 and 70 dB are chosen if not stated otherwise. Duration of both

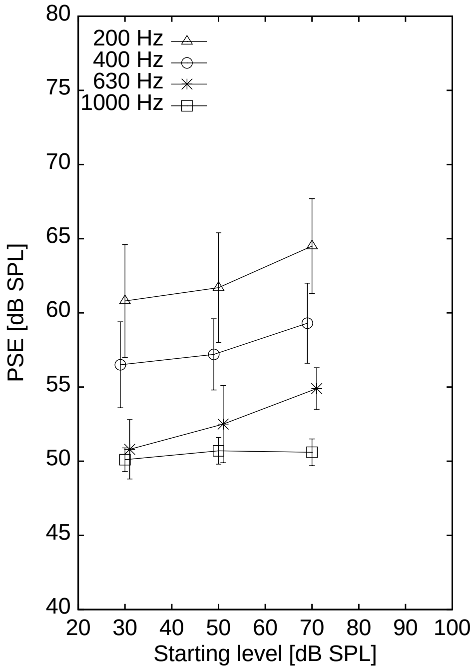

Figure 4.1.: *Mean PSE and standard deviation for 4 test-tone frequencies averaged across 6 listeners. Initial is step-size 8 dB. The abscissa indicates the starting levels for the adaptive procedure. The level of the reference is 50 dB.*

signals is 1 s with a silence interval of 500 ms between signals. Signals are gated with 50 ms Hanning ramps on and off. Test- and reference-signal are presented in random order. The same 6 subjects who performed the experiments in the previous chapter also participated in the experiments in this chapter. All subjects repeated each experiment four times. The different starting levels are presented in random order.

4.3. Influence of the starting level

In this experiment the influence of different starting levels is measured for the adaptive procedure. The results using an initial step-size of 8 dB are shown in Fig. 4.1 (and Table A.2 in the Appendix). Mean PSE and standard deviation for the 4 test-tone frequencies are indicated by different symbols: 200 Hz (△), 400 Hz (○), 630 Hz (×) and 1000 Hz (□), respectively. Mean PSE is shown as a function of the starting level.
An increasing PSE with increasing starting level is observed for all fre-

quencies except for 1000 Hz. The influence of the starting level is largest (4.2 dB) for the 630 Hz test-tone and statistical significant for the frequencies of 200, 400 and 630 Hz (Friedman Test, $p < 0.05$). For 200 and 400 Hz a change of starting level from 30 to 50 dB and from 50 to 70 dB is statistical significant (Wilcoxon Test, $p < 0.05$), whereas for 630 Hz only the change from 50 to 70 dB is significant (Wilcoxon Test, $p < 0.05$).

In contrast to the range effect reported by Gabriel [1996] from measurements with the method of constant stimuli an increasing effect with increasing spectral distance between test- and reference- signal is not observed. However, in the present study test-tone frequencies less or equal 630 Hz are employed whereas in the study of Gabriel the largest influence is observed for frequency distances from the 1 kHz reference smaller than 200 Hz, (see Fig. 2.2).

The interindividual standard deviation is 3.6 dB at 200 Hz and shows a decreasing tendency towards higher frequencies. For a test-tone frequency equal to the reference frequency a deviation of 0.9 dB is observed. Intraindividual standard deviation are almost independent of the frequency and are around 1.8 dB (not shown).

4.4. Influence of the initial step-size

In this experiment the influence of the initial step-size in the adaptive procedure is estimated. Besides the starting level this parameter has to be set by the investigator before the measurement. For a starting level far above or below the resulting PSE it takes a long time for the track to reach the PSE when a small step-size is used. When a large initial step-size is used the subject can reach after a few steps a level which is too loud or below threshold of hearing. The combined influence of initial step-size and starting level parameter on the result is investigated.

4.4.1. Initial step-size 16 dB

In this experiment the initial step-size is set to 16 dB and the influence of the starting level on the resulting equal loudness contours is systematically tested. Starting levels of 30, 50 and 70 dB SPL are used for all test-tone frequencies. In addition for the frequency of 200 Hz a starting level of 90 dB is presented.

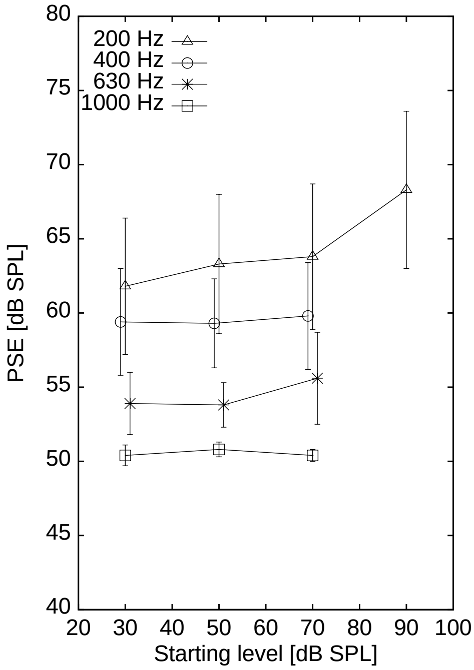

Figure 4.2.: *Mean PSE and standard deviation for 4 test-tone frequencies averaged across 6 listeners. Initial step-size is 16 dB. The abscissa indicates the starting levels for the adaptive procedure. The level of the reference is 50 dB.*

In Fig. 4.2 mean PSE values and standard deviations for the 4 test-tone frequencies are shown (for details see Table A.3 in the Appendix). The same symbols as in the previous figure are used for the different frequencies. The largest influence of the starting level in the range from 30 to 70 dB occurs for the test-tone frequency of 200 Hz. The effect amounts 2 dB. Thus, for the initial step-size of 16 dB, the effect is much smaller than for the initial step-size of 8 dB (see Table A.2 in the Appendix). For the other frequencies the influence of the starting level on the resulting PSE is smaller than 1.7 dB.

When the starting level is set to 90 dB for the 200 Hz test-tone, the resulting PSE increases by 4.6 dB in comparison to the starting level of 70 dB, even for the large initial step-size of 16 dB (see Fig. 4.2 and Table A.3 in the Appendix).

The influence of the starting level is only for the test frequency of 200 Hz significant when regarding the starting-level of 90 dB (Friedman Test, $p < 0.05$). For 200 Hz the change of the starting level from 30 to 50 dB and from 70 to 90 dB is statistical significant (Wilcoxon Test, $p < 0.05$), whereas the change from 50 to 70 dB is not significant (Wilcoxon Test, $p < 0.05$).

The interindividual standard deviations are 4.8 dB at 200 Hz and decrease towards higher frequencies, i.e. 3.4 dB at 400 Hz, 2.2 dB at 630 Hz and 0.6 dB at 1 kHz. Intraindividual standard deviation (not shown) are about 1.8 dB and almost independent of the frequency. The minimum intraindividual standard deviation is obtained at 1 kHz (1.3 dB) and the maximum at 630 Hz (2.2 dB).

One disadvantage of the large initial step-size is that it is only possible to use it for equal- loudness level contours which lie in the middle of the dynamic range. For the higher contours the test-tone can reach a level which is to loud, whereas at the lower contours the test-tone can already get below the threshold of hearing after one step.

4.4.2. Initial step-size 4 dB

Since the experiment with the initial step-size of 16 dB is performed after the experiment with the initial step-size of 8 dB, the reduced effect of the starting level could be a consequence of training. In order to test this hypothesis, this experiment is performed with an initial step-size of 4 dB, which is smaller than used in Exp.4.3. All other parameters are the same as in the previous experiment. If the effect of the starting level is larger than in Exp. 4.3 the hypothesis that training can reduce the influence of the starting level effect can be rejected.

The influence of the starting level using the adaptive procedure with an initial step-site of 4 dB is shown in Fig. 4.3 (and Table A.4 in the Ap-

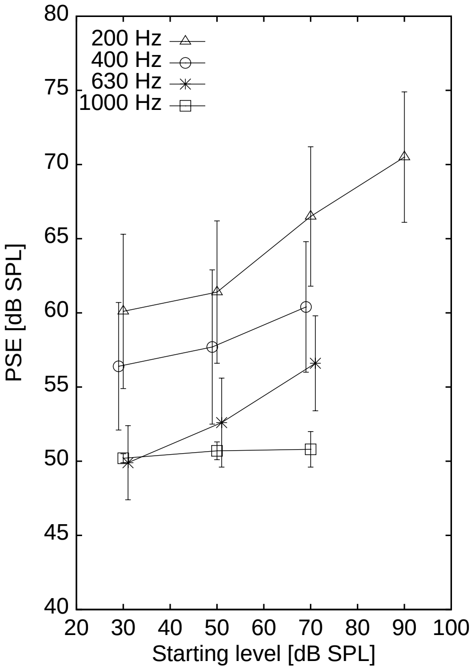

Figure 4.3.: *Mean PSE and standard deviation for 4 test-tone frequencies averaged across 6 listeners. Initial step-size is 4 dB. The abscissa indicates the starting levels for the adaptive procedure. The level of the reference is 50 dB.*

pendix). The difference between the resulting PSE for starting level of 30 and 70 dB amounts up to 6.6 dB for the test-tone with 630 Hz. The influence of the starting level is statistical significant for the frequencies of 200, 400 and 630 Hz (Friedman Test, $p < 0.05$). Except for the change in starting level from 30 to 50 dB for the 400 Hz signal, all other changes of starting level are statistical significant (Wilcoxon Test, $p < 0.05$). The interindividual standard deviations are 4.9 dB at 200 Hz with a decrease towards higher frequencies, i.e. 4.6 dB at 400 Hz, 2.9 dB at 630 Hz and 0.7 dB at 1 kHz. Intraindividual standard deviations are almost independent of the frequency and are around 1.8 dB, with a minimum of 1.4 dB at 1 kHz and a maximum of 1.9 dB at 630 Hz.

4.5. Peripheral versus central processing

In the previous experiments, signals are presented diotically to the subject, i.e. the same signal is presented on the left and the right ear. This presentation is more similar, compared to monaural presentation, to the free field condition which is required in ISO/TC43/WG1/N122 [1988]. It is not clear on which stage of the auditory pathway the interaction of methodical parameters, especially the starting level, with the loudness judgements takes place. In this experiment it is investigated if the influence of the starting level can be attributed to a central or peripheral process.

In the literature (e.g. Lehnhardt [1996]) for example the acoustical reflex is more a central process. When presenting a loud sound to one ear, a change of the compliance of the tympani membrane which is due to a contraction of the muscle in the middle ear, can be measured in the ear where the noise is presented as well as in the other ear which is in silence. An example for a peripheral process are otoacoustic emission (e.g. Mauermann et al.[2000]).

Marks [1996] found in magnitude scaling experiments as well as in direct comparison that perception of loudness was contingent on the distribution of tonal stimuli varying in sound frequency as well as SPL: When a low-frequency signal is presented at low level and a high-frequency signal at a high level, loudness of the low frequency signal is found to be large relative to that for the high frequency signal. When the association of level with frequency is reversed also the loudness relation is reversed. This phenomena is called recalibration. In order to decide if this is a peripheral or central process Marks performed several binaural experiments to investigate the influence of one ear information on the processing on the other ear. Following this idea in this section an experiment is performed in order to investigate if the bias effect is more a central or peripheral process.

4.5.1. Method

As in the previous experiments an adaptive 2-AFC procedure is used to test if the effect of starting level is due to a peripheral or a central process. In contrast to the diotic condition the test-tone is first presented on the left ear. When the step size is reduced to 2 dB after 4 turning points the

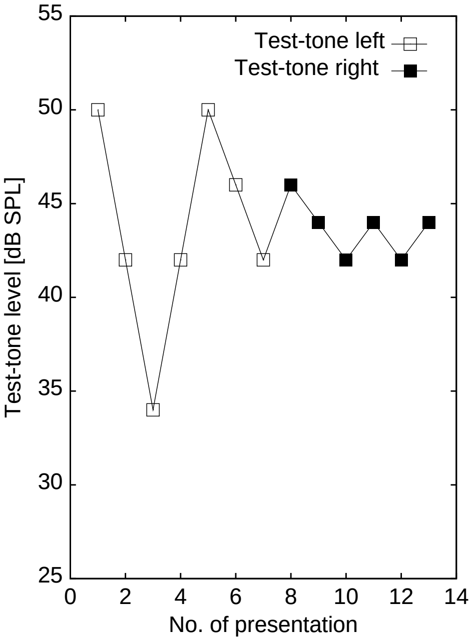

Figure 4.4.: *Example for the presentation of one test frequency in the adaptive procedure presentation. The reference-tone in this condition is always on the right ear. Abscissa: starting level for the adaptive procedure. Symbols indicate on which ear the test-tone is presented Test-tone left (□), Test-tone right (■).*

test-tone is presented on the right ear. The reference in this condition is always on the right ear. A simple adaptive procedure is used because a large influence of the starting level is desired in this paradigm (see Fig. 4.1 and Table A.2 in the Appendix). An example of a run is shown in Fig. 4.4.

4.5.2. Results

The influence of the starting level is similar as in the diotic experiment (Table 4.1 vs. Table A.2 in the Appendix). For all frequencies the difference between two starting level is ≤ 1 dB. The largest difference occurs at the frequency of 200 Hz where the effect in this experiment is 3.8 dB. In the diotically experiment (see Table A.2 in the Appendix) the results are shifted by 2.8 dB.

f [Hz]	30 - 50 dB	50 - 70 dB
200	0.8	3.8
400	0.7	1.9
630	1.5	1.8
1000	-0.2	-0.9

Table 4.1.: *Differences between PSE values using starting levels, i.e. 30 vs 50 dB and 50 vs 70 dB for 4 test-tone frequencies. In this condition the reference-tone is presented on the right ear. Initial step-size 8 dB.*

f [Hz]	30 - 50 dB	50 - 70 dB
200	0.4	3.4
400	1.1	1.8
630	1.2	2.1
1000	0.7	-0.8

Table 4.2.: *Same as in Table 4.1, but in this condition the reference-tone is presented on the left ear.*

In a control condition the experiment for measuring the influence of the starting level is performed for the other ear, too (Table 4.2). The reference-tone in this condition is always on the left ear. The test-tone starts on the right ear, when the step size is reduced to two dB after 4 reversals the test-tone is presented on the left ear, too. In average data over six subjects differ $\leq$ 1 dB from the results obtained in the last experiment.

Comparing the results from these experiments with the one obtained in the diotic condition one can see that the influence of the starting level on the resulting PSE is of the same amount. Thus, as a first assumption one could conclude that the effect of adapting to a starting level is not restricted to a peripheral process in the ear. Instead results indicate that it is more a centrally organized process.

4.6. Discussion

In this chapter is it shown that the influence of the starting level in an adaptive procedure is a process which is more centrally organized. The influence of the starting level on the resulting PSE is measured for frequencies ranging from 200 to 1000 Hz. In comparison to the method of constant stimuli the results are less influenced by the experimental setup (e.g. presented test-tone level). When changing the starting level by 40 dB the PSE is shifted to a higher level of about 4.2 dB (Initial step-size 8 dB, see Exp. 4.3, Table A.2 in the Appendix).

A further reduction of the influence of the starting level can be achieved by employing larger initial step-size for the adaptive procedure. The influence is around 2 dB with an initial step-size of 16 dB for a change of 40 dB of the starting level (Exp. 4.4.1). But when starting at a level of 90 dB the resulting PSE is shifted by 4.6 dB, in comparison to the result obtained for an starting level of 70 dB. Thus, in the adaptive procedure a change of 60 dB of the starting level (from 30 to 90 dB) changes the resulting PSE by 6.6 dB (Table A.3). However, for a small initial step-size of 4 dB in the adaptive procedure results are strongly influenced by the starting level.

In comparison to the experiments with the larger initial step-size, the influence of the starting level is enlarged when employing an initial step-size of 4 dB. This experiment is performed after the experiments with initial step-size of 8 and 16 dB to test the hypothesis that the reduction of the influence of the starting level in Exp. 4.4.1 compared to Exp. 4.3 is due to training effects. Because the smallest initial step-size lead to results which are at most influenced by the choice of the starting level, the hypothesis that training is the reason for a reduction of the influence of the starting level can be rejected.

Simulations to test the influence of the initial step-size for the estimation of the threshold of hearing were done by Green et al. [1989]. Initial step-size was varied from 1 to 8 dB. The largest bias of around 6 dB was obtained with the smallest initial step-size. Larger initial step-size of 8 dB reduced the bias to 1 dB. The finding that due to a larger step-size the influence of the experimental setup can be reduced is in agreement with the data in this work.

The results obtained using an adaptive procedure are less influenced by the starting level compared to the results measured with the method of constant stimuli. Although a further decrease of the influence of the employed starting level can be achieved when using a large initial step size, a shift in starting level of 60 dB can yield a PSE that differs by about 6.6 dB.

One disadvantage of the large initial step-size is that it is only practicable for the determination of equal-loudness level contours which lie in the middle of the dynamic range. For the higher contours the test-tone can reach a level which is to loud, whereas at the lower contours the test-tone can get already under the threshold of hearing after a few steps.

Therefore, a more complex method should be used which might lead to less biased results such as an adaptive procedure with interleaved tracks.

5. Interleaved procedure

5.1. Introduction

As shown in the previous chapter, the choice of the starting level has an influence on the point of subjective equality (PSE) when a simple adaptive procedure (with one track) is used. The effect can be reduced by choosing a large initial step size. Nevertheless, also with the largest investigated step-size of 16 dB the difference between the $PSEs$ obtained with two starting levels that are 60 dB apart from each other still amounts to 6.6 dB. Since this difference is larger than the interindividual variation, it is desirable to reduce further the influence of methodical factors.

A promising variation of an adaptive procedure was proposed by Jestaedt [1980]. He argued that response bias could be avoided if instead of one track (as in the simple adaptive procedure) two tracks with different starting levels should be interleaved in each experimental run. He recommended this procedure especially for measurements of subjective impression such as lateralization or loudness. In recent studies concerning loudness, procedures were frequently used that were based on Jesteadt ideas (e.g. Florentine et al. [1996] and Verhey [1999]).

Although not all experimental details are comparable in those studies, the authors were convinced about the improved reliability of data obtained with interleaved adaptive tracks. However, most of the studies investigated loudness as a function of other parameters than frequency such as duration or bandwidth. In addition only a few remarks were made by the

authors about the influence of methodical factor on the results.
In the present study the effect of varying methodical parameters on the measured PSE for tonal signals is investigated for adaptive procedures with interleaved tracks. In the first experiment, two tracks with the same frequency are interleaved. This procedure is close to the procedure proposed by Jesteadt. In the second experiment, four tracks with different test-signals for each track are interleaved. This procedure is comparable to the procedure used by Verhey [1999] to measure spectral loudness summation. In both experiments the effect of the starting level is quantified by choosing two different starting level ranges. In the third experiment mechanism underlying the methodical effects are investigated.

5.2. Interleaved procedure with two tracks

5.2.1. Methods

The same subjects as in the previous experiments participate in the experiments. The test-tone frequencies are 200, 400 and 630 Hz. The frequency of the reference is always 1000 Hz at a level of 50 dB. The test-tone and the reference-tone are presented in random order. The duration for both signals is 1 s with a silence interval of 500 ms between the signals. Test- and reference-signal are presented via headphones Stax SR λ, (for further details see section 3.1.1). Each subject repeated the experiment four times.
The employed method is an adaptive 2-AFC procedure. The initial step-size is 8 dB. After each reversal point, the step-size is halved and the minimum step-size is 2 dB. In contrast to the previous experiments two interleaved tracks are presented (see Fig. 5.1). In Exp. 5.1 one low and one high level condition is employed. In the low level condition the starting levels are 30 and 70 dB. In the high level condition starting levels of 50 and 70 dB are used. Thus, the difference between starting levels remains constant, whereas the absolute position of the level is shifted by 20 dB.

Results

For each frequency, the PSE values obtained with two tracks with different starting levels differ less than 1 dB (not shown). This finding holds for both conditions. The tracks converge almost independently of the employed starting level to the same point. Therefore only averaged results over the two starting levels of the interleaved tracks are shown in Table 5.1. A shift of the starting level has still a small effect on the measured PSE for 400 and 630 Hz. The difference between these two conditions is smaller than 1 dB.

However, for a test-tone frequency of 200 Hz values with the higher starting levels are shifted up to 3.3 dB compared to results with low starting levels. Note that this effect occurs only for the largest spectral distances between test- and reference-tone and only for this frequency the effect is significant (Wilcoxon Test, $p < 0.05$).

f [Hz]	200	400	630
Interleaved low (30, 70)	63.5	60.1	55.4
Interleaved high (50, 90)	66.8	60.8	55.9

Table 5.1.: *Mean PSE for the the low and high level condition. Data are for 3 frequencies averaged across 6 subjects. Standard deviation between the tracks is around 1 dB.*

The results obtained with the adaptive procedure with two interleaved tracks can still be influenced by the employed starting levels. Subjects might have the tendency to compare not only the reference with the test-tone, but to adjust test-tones of the two track procedure to give the same loudness sensation.

A more complex procedure is needed, to avoid comparisons between the two tracks. This might although help to reduce the influence of the starting level to an amount which is indistinguishable from the exactness of the measurement procedure, determined by the minimum step-size and of individual reproducibility.

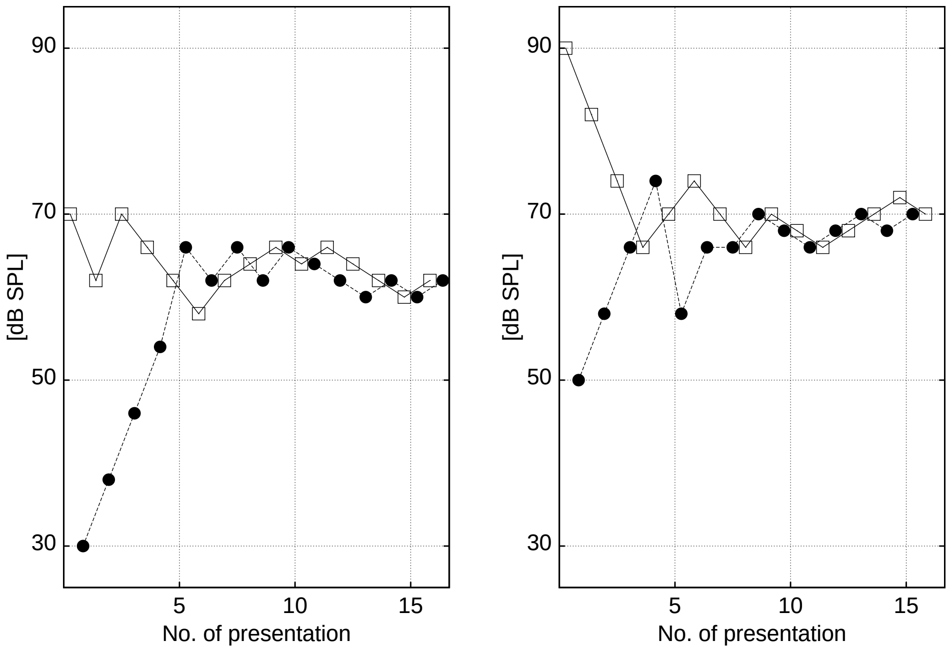

Figure 5.1.: *Left panel: Example for the presentation of one test frequency in the adaptive interleaved procedure with 2 tracks. Abscissa: No. of presentation. Ordinate: starting level of the adaptive procedure. Initial step-size is 8 dB. Symbols indicate two different starting levels in this case 30 dB (●) and 70 dB (□) for the test-tone. The test-tone frequency is the same for both tracks. Right panel: Same as in left panel, but for starting levels of 50 (●) and 90 dB (□).*

5.3. Interleaved procedure with four tracks

The data obtained with two interleaved tracks indicate that the results are still influenced by the starting level. In this section, a more complex interleaved procedure is tested. As in the previous experiment the point of subjective equality (PSE) is measured for test-tone frequencies of 200, 400, 630 and 1000 Hz. The reference frequency is 1000 Hz at a level of 50 dB SPL. In contrast to the previous experiment, for each track a different test frequency is used. This is done to focus subjects attention to each trial and not to follow the course of the test-tone levels in each adaptive track. As in the previous experiment two conditions are investigated. In a high level condition starting level of the tracks are 50, 70 and 90 dB, whereas in the low level condition starting levels of 30, 50 and 70 dB are used. Although it is shown in the previous chapter that a large step-size reduces the effect of the starting level, the disadvantages (e.g. reaching level which are too loud or below threshold after a few steps) dominate and a medium step-size of 8 dB is used.

5.3.1. Method

Subjects and measurement setup is the same as in the last experiments, with the following exception.

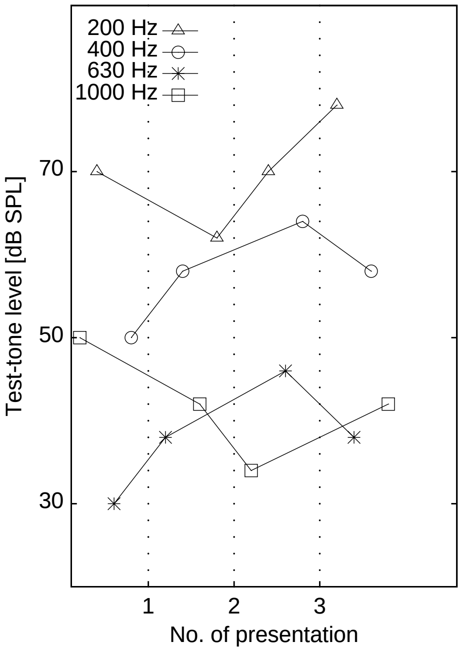

Figure 5.2.: *Example for the first four presentations of the test-signals in the adaptive interleaved procedure for the condition with low level. Ordinate starting level of the adaptive procedure. Symbols indicate four different test-tone frequencies.*

In contrast to the previous experiments four interleaved tracks with different starting-levels are presented (see Fig. 5.2). Since the time for one measurement increases due to the amount of tracks the paradigm is limited to four tracks.

In the low level condition the starting levels are 30, 50 and 70 dB and in the high level condition 50, 70 and 90 dB, thus, the difference between starting levels and the level range remains constant whereas the absolute position of the range is shifted.

5.3.2. Results

In Fig. 5.3 results for both conditions are shown. In the left figure starting levels are 30, 50 and 70 dB (condition with low levels) and in the right figure starting levels are 50, 70 and 90 dB (condition with high levels). The influence of the initial starting level for both experiments is smaller than 2.3 dB. The largest influence of the starting level can be seen in the high level condition. A change of starting level from 50 to 70 dB at a test-tone frequency of 200 Hz shifts the result around 2.3 dB. The effect is almost independent of the test-tone frequency except for the 1 kHz test-tone, where the effect is < 0.5 dB.

Although the influence is not significant across the conditions, a change in starting level from 50 to 70 and from 70 to 90 dB is statistical significant for 200 Hz, and a change 70 to 90 dB is significant for 400 Hz (Wilcoxon Test, $p< 0.05$). Comparing the results for the low and the high level conditions for the same starting level the largest difference is at 200 Hz for a starting level of 70 dB, whereas the PSE in the high level condition is 1.8 dB higher. The same difference between the two conditions is at 400 Hz and a starting level of 50 dB, again the PSE measured in the high level condition is 1.8 dB higher. Only the effect for 400 Hz is statistical significant (Wilcoxon Test, $p< 0.05$).

All other differences are smaller than 1 dB and not systematically higher for the high level condition. Results for all frequencies and both level ranges are shown in Fig. 5.3 and Table 5.2 and Table 5.3.

Comparing the mean data from both experiments in the low and in the high level condition, a 40 dB shift of the starting levels has only a small

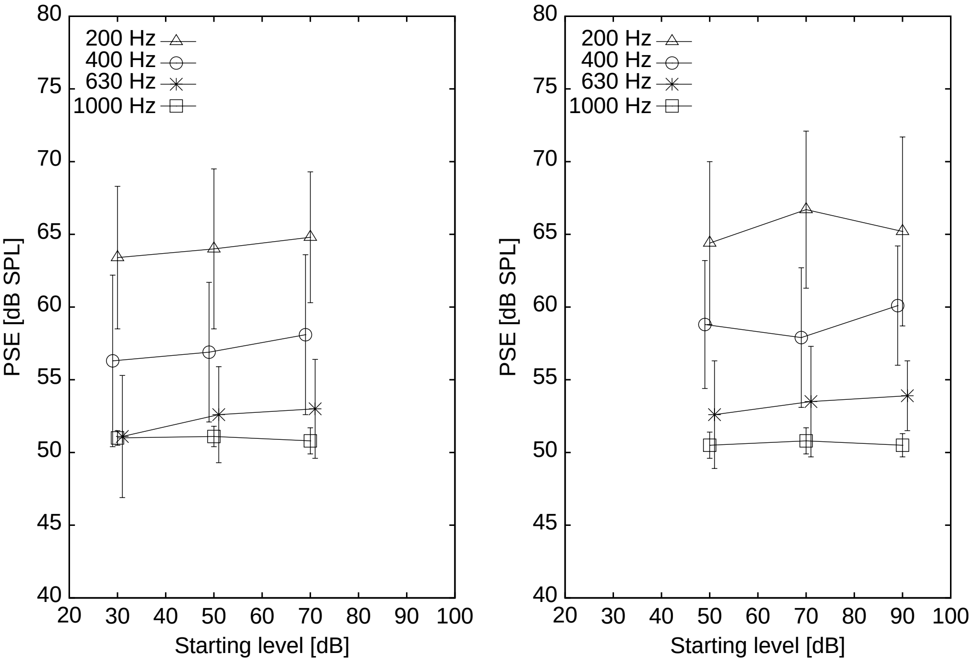

Figure 5.3.: *Left panel: Mean points of subjective equality (PSE) and standard deviation for 4 test-tone frequencies averaged across 6 subjects. Ordinate: PSE. Abscissa: starting level of the adaptive procedure (30, 50 and 70 dB). Symbols indicate four different test-frequencies 200 (△), 400 (○), 630 (*) and 1000 Hz (□). Right panel: Same as in left panel, but starting levels are 50, 70 and 90 dB.*

influence on the results (< 2 dB).

In comparison to data obtained with the simple adaptive procedures (see Fig. 4.1 to Fig. 4.3), interindividual standard deviations are enlarged. In comparison to the standard deviation obtained with the simple adaptive procedure the standard deviation with interleaved tracks is by 2.6 dB (200 Hz), 3 dB (400 Hz) and 2.1 dB (630 Hz) larger. At 1 kHz the standard deviation is constant ≤ 1.1 dB for the procedures.

Intraindividual standard deviation is almost independent of the employed procedure and always ≤ 2.2 dB and in most cases standard deviation is ≈ 1.5 dB, whereas no clear dependency can be seen from the employed procedure and from the the test-tone frequency.

f [Hz]	30 - 50 dB	50 - 70 dB
200	0.58	0.75
400	0.63	1.21
630	1.92	-0.04
1000	0.21	-0.42

Table 5.2.: *Difference in resulting PSE between starting level 30 and 50 dB and starting level 50 and 70 dB.*

f [Hz]	50 - 70 dB	70 - 90 dB
200	2.29	-1.38
400	-0.58	1.79
630	0.90	-0.40
1000	0.29	-0.25

Table 5.3.: *Difference in resulting PSE between starting level 50 and 70 dB and starting level 70 and 90 dB.*

5.4. Summary and Discussion

The effect of the employed starting level is obviously reduced in the experiment using four interleaved tracks in comparison to data obtained using the method of constant stimuli or the simple adaptive procedures with only two tracks.

The data measured with an interleaved adaptive 2-AFC procedure are nearly not influenced by the experimental setup. Respectively, the influence of the starting level is of same amount as the individual reproducibility indicated by the interindividual standard deviation.

Besides the effect that it is not possible for the subjects to follow the course of the tracks in the interleaved procedure with 4 tracks (and therefore the influence of the starting level is reduced) a disadvantage can be observed: the standard deviation is enlarged in comparison to the results obtained with the simple adaptive procedures.

Although the results are not strongly influenced by the starting level a good choice of the starting level, which has to be done in advance by the investigator, can reduce the influence of the starting level. Good choice can be interpreted as not too far away from an expected value. As shown in the last experiments there is a range where the starting point has nearly no influence on the result as long as it is in a sensible range. Instead of choosing a starting level and indicating the difference between the starting level an the reference in dB, it is more adequate to look for the difference to the corresponding PSE or a loudness value. Since 70 dB SPL starting level at 200 Hz leads to a different loudness impression as a 70 dB SPL starting level at 630 Hz. While in the first case 70 dB leads to an almost equal-loud impression in comparison to a 50 dB reference-tone, in the second case it is much louder.

5.5. Individual differences and threshold of hearing

Measured PSE values and data from the literature show considerable differences especially at low levels and for test-tone frequencies far away from the reference frequency (see Fig. 5.4). The effect that intraindividual variance is larger than interindividual variance is reported by several authors e.g. Gabriel [1996], Betke [1991] and Suzuki et al. [1989].

Taking a closer look to individual data one can see different groups of subjects, which repeat measurements with high reproducibility but differ in their results from other subjects. A characteristic common shape of all $ELLC$ is found for each single subject. For example, if the PSE is low at 30 phon for one subjects for a certain frequency it is also low for the 50 phon contour at this frequency. Data from Betke [1991] show interindividual standard deviation about 5.3 dB (13, subjects 50 Hz, 40 phon). Gabriel [1996] found deviations about 6.8 dB (6, subjects 500 Hz, 30 phon) for $ELLCs$.

These standard deviations are of the same amount as those for the threshold of hearing as found by Betke. Similar standard deviations were also observed by Watanabe & Møller [1990]. For $ELLC$ 7.1 dB are obtained (12 subjects, 63 Hz, 20 phon) and 6.5 dB for the threshold of hearing (12 subjects, 63 Hz). Although these large differences are described in many papers only in few papers an explanation is proposed to account for these interindividual differences.

In the investigation of Gabriel [1996] $ELLCs$ for different listeners are compared with individual head related transfer functions ($HRTF$). The idea was that equal-loudness level contours at low levels are of similar shape as the threshold of hearing. Gabriel showed that individual threshold can be, at least partly, attributed to the individual $HRTF$. In Garbiel's work this is shown for 3 subjects for frequencies below 1 kHz. In a work of Takeshima et al. [1994] a significant correlation between $HRTF$ and the threshold of hearing is found for frequencies between 1 and 8 kHz,

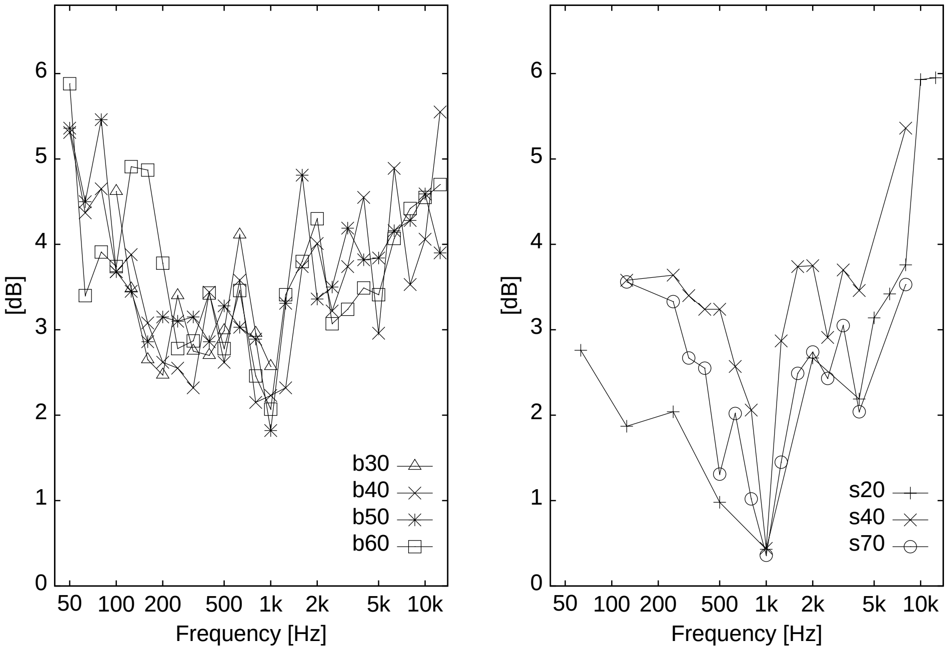

Figure 5.4.: *Left panel: Interindividual standard deviation, data from Betke [1991] for 30 (△), 40 (×) 50 (∗) and 60 (□) phon. Right panel: Interindividual standard deviation data from Suzuki [1989] standard deviation for 20(+), 40(×) and 70(○) phon.*

whereas below 1 kHz the results are not significant. At least parts of individual differences can be attributed to the individual $HRTFs$. This fine structure will probably have influence also in $ELLCs$.

From Fig. 5.5 one can see that the shape of the threshold is almost similar to the shape of the lower $ELLC$. Subtracting the threshold of hearing from $ELLCs$ should lead on average, at least for the lower $ELLC$, to a straight line. The results are shown in Fig. 5.6. Since the resulting curves for the data of ISO 226 are almost a straight line, in the data from Betke [1991] deviations are observable. The reason for these small minima and maxima can not be explained. An investigation for individual is not possible since these data are not available.

From the literature (e.g. Cohen [1982], Schloth [1983], Long [1984] and Long et al. [1988]) it is known that threshold of hearing is not just a

smooth change of level vs. frequency but has a fine structure. Threshold shows a consistent pattern of minima and maxima. These pattern are characteristic for each subject and shape of the fine structure is remarkably constant over a long time (Long [1984]).
The findings in this work and earlier published paper are used to explain intraindividual differences by taking into account the individual threshold of hearing for every subject at the distinct test-tone frequencies (Reckhardt et al. [1998], Reckhardt et al. [1999]).
In order to test this hypothesis threshold of hearing is determined for the 4 test-tone frequencies.

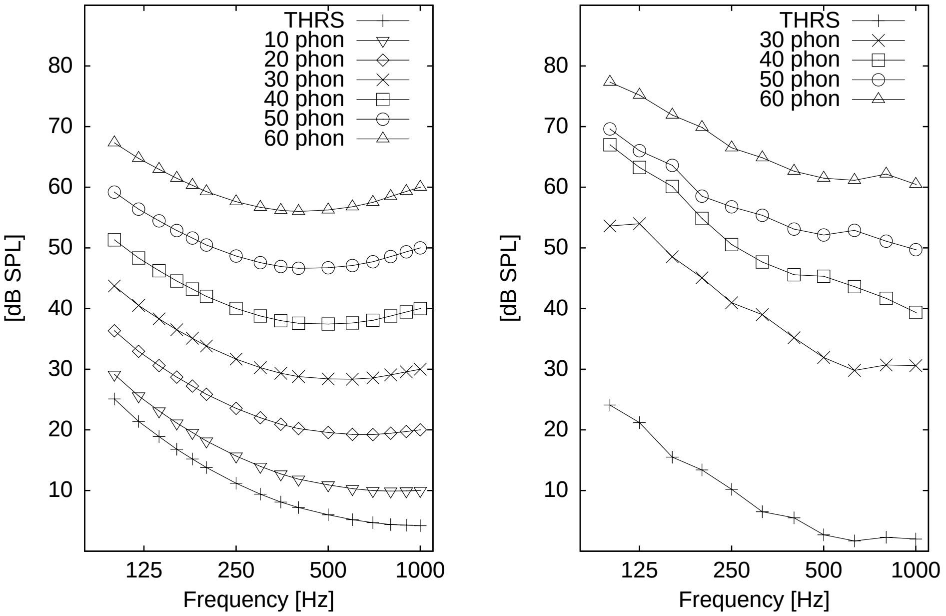

Figure 5.5.: *Left: Threshold of hearing and equal loudness level contours defined in ISO 226 in the range from 100 to 1000 Hz and from 10 to 60 phon in steps of 10 phon. Right: Threshold of hearing and 30, 40, 50 and 60 phon contour from Betke [1991] in the range from 100 to 1000 Hz.*

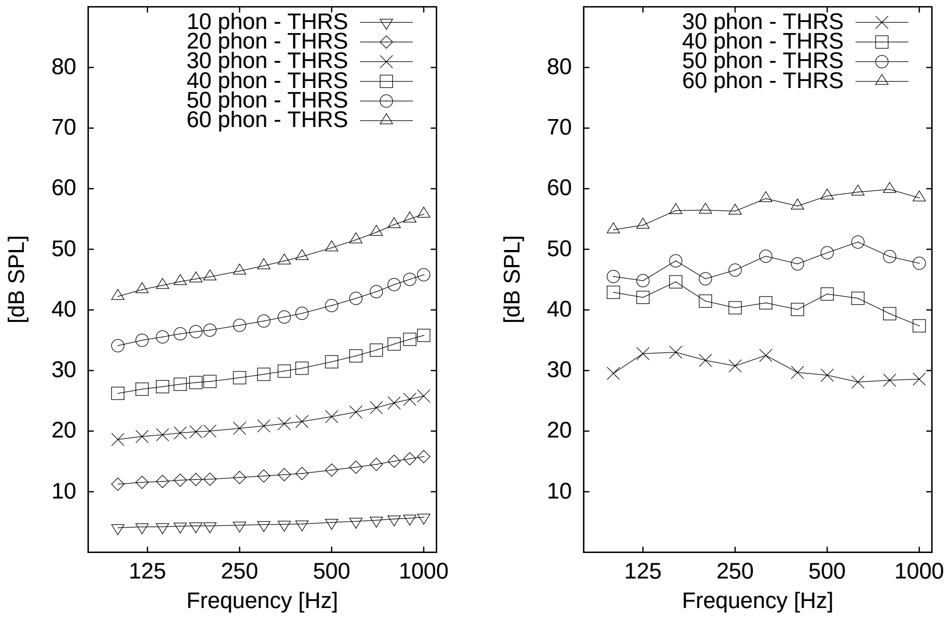

Figure 5.6.: *Left panel: Same as Fig. 5.5(left panel) but threshold of hearing is subtracted from equal-loudness leve contours. Right panel: Same as in Fig. 5.5(right panel) but threshold determined by Betke [1991] is subtracted from 30, 40, 50 and 60 phon contour.*

5.5.1. Measurement procedure

For the determination of the threshold of hearing a 3 AFC 1 up 2 down procedure is used. This procedure estimates the 70.7 % point of the psychometric function (Levitt [1971]). Three intervals are presented which are indicated by a small light. Duration of the signals is 1 s with a separation of 500 ms between the signals. Signals are gated with 50 ms Hanning ramps on and off. In one randomly chosen interval the test-tone is presented. Subjects' task is to indicate which of the three intervals contains the test-tone. The response is given by a keyboard by pressing a number between 1 and 3 corresponding to the signal position. On the computer screen a feedback is given if the answer is correct or not and then the next

trial is presented.
The initial step-size is 8 dB. Depending on the answer of the subject level of the signal is in- or decreased. After each reversal the step-size is halved. The minimum step-size is 1 dB. After four reversals with minimum step-size the measurement is finished. All experiments are repeated four times. The threshold of hearing is calculated as the median of these four last reversals. The individual threshold is determined for each subject with the measurement setup described in chapter 3.1.1. Threshold of hearing is determined for the same six subject as in the previous experiments. Test-frequencies are 200, 400, 630 and 1000 Hz. Every measurement is repeated 4 times.

5.5.2. Results

Results are shown in Fig. 5.7 (left panel). One can see large differences between subjects, but intraindivudual variance is rather small. The upper graph shows individual PSE for a frequency of 200 Hz measured with an adaptive 2-AFC procedure with interleaved tracks (see Exp. 5.3). Data are shown for a starting level of 70 dB.
The shape of both contours looks quite similar. Subjects with a relatively high threshold of hearing have a high PSE in the loudness measurements, too. In the right panel of Fig. 5.7 individual PSE and values subtracted by individual threshold of hearing are shown. Taken into account the individual threshold of hearing standard deviation between subjects is almost halved. The obtained contours can be interpreted as *equal-loudness level contours sensation level* ($ELLC - SL$).
When subtracting individual threshold from individual PSE, measured with the interleaved procedure, interindividual standard deviation is reduced for 200 Hz from 6.2 dB to 2.6 dB, 400 Hz from 6.4 dB to 4.3 dB whereas for 630 Hz standard deviation remains almost unchanged at about 4 dB. These values are averaged data for the three employed starting lev-

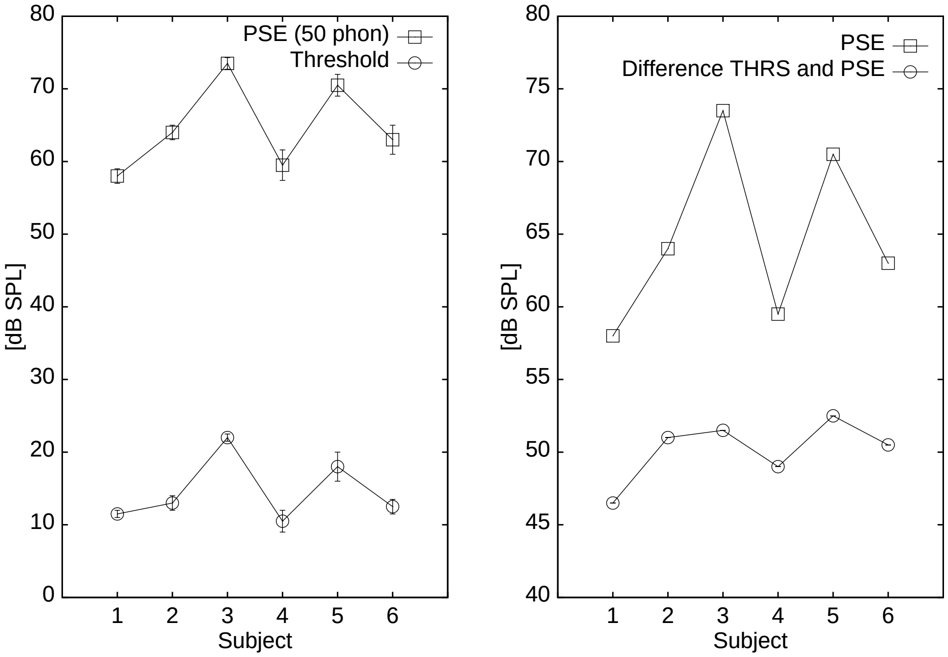

Figure 5.7.: *Left panel: Individual PSE (starting level 70 dB, test-tone frequency 200 Hz) and individual threshold of hearing for six subjects. Right: Individual PSE and values subtracted by individual threshold of hearing.*

els.

The influence of the threshold on the $ELLC$ is probably restricted to low and medium levels because for higher levels the $ELLC$ are getting smooth and for the highest level the shape is almost a straight line (see Fig.1.1 and Fig 5.5) and differs from the threshold of hearing.

In order to test if the influence of the threshold is stronger for lower $ELLC$, threshold data are subtracted from the 30 phone contour determined in section 3.2.5 .

Subtracting the threshold from individual PSE reduces standard deviation for 200 Hz from 4.4 dB to 2.2 dB, for 400 Hz from 3.6 dB to 3.1 dB, whereas for 630 Hz the standard deviation is enlarged from 3.5 dB to 4.1 dB. An explanation for the smaller influence might be that the $ELLCs$ for 30 phone are not obtained with an interleaved procedure, but with a simple adaptive method and are therefore biased. Standard deviations for the measured data with the simple adaptive procedure are smaller than those obtained with the interleaved procedure.

These influence of the threshold on equal-loudness level contours was confirmed for low frequencies in a study of Bellmann et al. [1999]. The 60 phon curve was determined in the range from 16 to 160 Hz and the threshold of hearing from 40 to 160 Hz for six subjects. The procedures for the determination of the equal-loudness level contour was the same as in Exp. 5.3 and the measurement of the threshold of hearing was the same as in Exp. 5.5.1. Taking into account the individual threshold variance was reduced by 6 to 10 dB depending on the frequency.

In the same way as Reckhardt et al. [1998] and Reckhardt et al. [1999] (where parts of the data of this section were presented), Mauermann et al. [2000] investigated the relation between the fine structure of the threshold of hearing and *ELLC*. In this recent paper the signal duration was 250 ms and the frequency range was limited from 1400 to 2000 Hz. The lower limit was set due to unwanted noises of the subject (e.g. breathing, small movements). However, the authors argued that the effect of the fine structure of the threshold of hearing and of the *ELLCs* is most probably not restricted to this region.

The threshold of hearing was determined by Mauermann et al. [2000] with an adaptive 3 alternative choice 1-up 2-down procedure in steps of 25 Hz. This is similar to the procedure used in the present study to determine the threshold in Exp. 5.5.1. Variations in the threshold of up to 10 dB for a small frequency change of around 50 Hz are observed.

Equal-loudness level contours were measured with the same interleaved adaptive 2-AFC procedure used in the present investigation and in the paper from Verhey [1999] (see Exp. 5.3). Some of the data from Mauermann were measured with a direct comparison experiment, but differences between the procedures were not discussed.

Equal-loudness level contours were measured in 5 phon steps. The fine structure for the equal-loudness level contours at low level was almost

similar to the one for the threshold of hearing. For higher level the curves were smoother and for level higher 60 phon the shape was almost a straight line.
These findings support the proposal taking the individual threshold into account in an $ELLC$ measurement.

5.6. Conclusion

The obtained contours when the threshold is subtracted from equal-loudness level contours can be interpreted as *equal-loudness level contours sensation level level* ($ELLC - SL$) (see Fig. 5.6 and Fig. 5.7(left panel)). Since subjects threshold's have a fine structure, differences between subjects in the determination of equal-loudness level contours reflect these individual differences in the threshold of hearing. At least for the contours at low levels the influence of the threshold fine-structure can be eliminated from the $ELLC$ by calculating an $ELLC - SL$.

6. General Conclusion

The influence of experimental parameters on the results in experiments concerning equal-loudness level contours is investigated. A model which is proposed by Gabriel [1996] (and fitted to data which are obtained under the same experimental conditions as used by Betke [1991] and Fastl et al. [1990]) is applied to data from the literature. In the model differences between results are attributed to the employed experimental setup. The model takes in particular the absolute position of the employed test-tone level range influences the resulting point of subjective equality (PSE) into account. Gabriel [1996] adjusted the model parameters to minimize the differences between data which are determined under comparable conditions (e.g. by Betke and Fastl for the 30 phone equal-loudness level contours ($ELLC$)). The model with this set of parameters fails to minimize differences between data from literature for other $ELLCs$. Even for the original data for the 30 phon contour the aim to minimize differences can only partly be achieved. Although for some frequencies the calculated curves seem to differ less, for others the difference is increased.
A modification of the determination of model parameters is investigated. Unreasonable data can be avoided by modifying the calculation of the REL value, which reflects the presented test-tone levels in the measurement procedure. Instead of calculating REL, as proposed by Gabriel [1996], from the maximum and minimum of the presented level with a weighting of 80% to 20%, a substitution by using a loudness calculation with the N_{35} percentile avoids unreasonable low values. But the 'correct' (i.e. 'absolute')

$ELLC$-values remains still unclear. Since only some differences could be explained by the model, the requirements and limitations of the parameters involved are investigated. Is is shown that for some frequencies it is impossible to minimize the differences between $ELLC$ from different studies with this model.

Considering the findings of Gabriel and respecting the recommendations from ISO/ TC43/ WG1/N122 [1988], an adaptive procedure is proposed for the determination of $ELLC$. First results obtained with the method of constant stimuli and with the adaptive procedure are compared. The influence of the measurement setup on the results is smaller in comparison to the results obtained with the method of constant stimuli.

The parameters which have to be set by the investigator in the adaptive procedure are systematically varied. The influence on the resulting PSE is quantified and it is shown that a large initial step-size in the adaptive procedure minimizes the influence of the starting level. While a large initial step-size has the disadvantage that the test-tone can get too loud or below threshold of hearing after a few steps, a moderate initial step-size of 8 dB is proposed. It is shown that in a simple adaptive procedure the starting level can still influence the result. Even with two interleaved tracks the results are biased. Instead of presenting one adaptive track, at least 4 frequencies with different starting levels have to be randomly presented in an interleaved procedure in order to minimize the influence of the experimental parameters. The influence of the starting level on the resulting PSE can be reduced to the same order of individual reproducibility. Therefore, the proposed method is insensitive enough against the experimental setup and can be used for the determination of equal-loudness level contours.

Verhey [1999] reported that the influence of the starting level almost vanishes for shorter signals (10 ms). Maybe due to the fact that these short signals do not produce a hearing sensation of pitch, the task is equal difficult for all frequencies. Heldmann [1994] claimed that for the impression of pitch 35 ms are necessary (frequency > 100 Hz). In might be concluded that bias effect in the determination of $ELLC$ can be reduced by taking

shorter signals. From Gabriel[1996] it is known that the bias is largest for the largest differences between test and reference-tone. This is because for subjects it is most difficult to compare tones which differ in pitch.
The reason for the bias remains unclear. At least from one experiment it is concluded that the process can be attributed to a more central mechanism in the processing of acoustic sensation. Although it is clear that it is a difficult task for the subject comparing two different stimuli, measurements are repeated with high reproducibility. But the variance between subjects is rather large. All the data from the literature show that the variance between individuals (intraindividual variance) is much smaller than differences between different subjects (interindividual variance). This variance can partly be minimized when taking into account the individual threshold of hearing. When subtracting the individual threshold of hearing from the individual PSE for each frequency, at least at low levels and low frequencies the standard deviation is reduced about a factor of two. The obtained equal-loudness level contours can be interpreted as *equal-loudness level contours sensation level level* $(ELLC - SL)$.

The proposed measurement setup for the determination of new $ELLCs$ should be restricted to the proposed schemes in order to avoid biasing the results by the employed procedures. The results of this investigation will help to determine comparable loudness level contours in different laboratories under free-field conditions in order to be used for a revision of the contours for a new standard.

A. Appendix

A.1. Comparison of data from the literature

A comparison of papers from Møller & Andresen [1984] and Watanabe & Møller [1990] is given in Table A.1:

Phon	Frequency	(1)	(2)	(3)
20	63	58.0	60.8	65.7
40	63	71.7	75.9	81.4
60	63	82.8	-	93.3

Table A.1.: *Comparison of data measured by: (1)= Møller & Andresen [1984] (adaptive procedure), (2)= Watanabe & Møller [1990](adaptive procedure), (3)= Watanabe & Møller [1990] method of limits.*

The main difference between the papers is that Møller & Andresen [1984] perform the experiments in an pressure chamber and signals have a duration of about 2 sec separated by a silent interval of around 1 s , whereas Watanabe & Møller [1990] use an unechoic chamber and signals have a duration of 1 s, separated by a 0.5 s silent intervall.

While it is unclear which is the reason for the large differences between data, for the model investigation in chapter 2 the newer data obtained with the adaptive procedure are used.

A.2. Original instruction displayed on the computer screen

WELCHES DER BEIDEN SIGNALE WAR LAUTER ?

Drücken Sie Taste $\boxed{1}$, wenn das erste Signal lauter war.

Drücken Sie Taste $\boxed{2}$, wenn das zweite Signal lauter war.

A.3. Instructions for the measurement procedure

Instructions for the measurement procedure (same as used by Gabriel [1996]).

Instruktionen zum Versuchsablauf

Bei diesem Versuch werden Ihnen jeweils zwei aufeinanderfolgende Töne über den Kopfhörer vorgespielt. Die Töne unterscheiden sich in der Tonhöhe und in der Lautstärke. Bitte beurteilen Sie nur die Lautstärke der Töne und beantworten Sie nach der Darbietung der beiden Töne die folgende Frage:
Welcher der beiden Töne war lauter?

Drücken Sie Taste 1 auf der Tastatur, wenn der erste Ton lauter war.
Drücken Sie Taste 2 auf der Tastatur, wenn der zweite Ton lauter war.

Wenn Sie Ihre Antwort abgegeben haben, werden Ihnen zwei neue Töne vorgespielt.
Vielen Dank für Ihre Mitarbeit!

Instructions for the measurement procedure

In this experiment you will be presented with two successive tones via headphone. The tone will differ in pitch and in loudness. Please, base your judgement only on the loudness and answer after the presentation of two tone the following question:
Which of the two tone was louder?

Press key 1 on the keyboard, if the first tone was louder.
Press key 2 on the keyboard, if the second tone was louder.

When you have made your judgement, you will be presented with a new pair of tones.
Thank you very much for your co-operation!

A.4. Data from Chapter 4

f [Hz]	30 - 50 dB	50 - 70 dB
200	0.9	2.8
400	0.7	2.0
630	1.8	2.4
1000	0.5	-0.1

Table A.2.: *Differences between PSE values using different starting levels, i.e. 30 vs. 50 dB and 50 vs. 70 dB for 4 test-tone frequencies. Initial step-size 8 dB.*

f [Hz]	30 - 50 dB	50 - 70 dB	70 - 90 dB
200	1.5	0.5	4.6
400	-0.1	0.5	-
630	-0.1	1.8	-
1000	0.3	-0.4	-

Table A.3.: *Same as in Table A.2, but initial step-size is 16 dB. In addition a starting level of 90 dB is used for 200 Hz.*

f [Hz]	30 - 50 dB	50 - 70 dB	70 - 90 dB
200	1.3	5.1	4.0
400	1.3	2.6	-
630	2.7	3.9	-
1000	0.5	0.1	-

Table A.4.: *Same as in Table A.3, but initial step-size is 4 dB.*

REFERENCES

Bellmann, M.A., Mellert, V., Reckhardt, C. and Remmers, H. 'Sound and Vibration at Low Frequencies', Psychophysics, Physiology and Models of Hearing, T.Dau, V.Hohmann, B.Kollmeier (Editors) Word Scientific,isbn 981-02-3741-3, pp. 117-120, **1999**

Berger, E.H. 'Re-exemination of low-frequency (50-1000 Hz) threshold of hearing in free and diffuse sound fields', J.Acoust.Soc.Am.**70**, pp. 1635-1645, **1981**

Betke, K. 'Hörschwellen und Kurven gleicher Pegellautstärke im ebenen Schallfeld', Dissertation, Universität Oldenburg, **1991**

Brinkmann, K., Vorländer, M., Fedtke, T. 'Re-determination of the threshold of hearing under free-field and diffuse field listening conditions', Acustica **80** pp. 453-462, **1994**

Buus, S., Müsch, H. 'On loudness at threshold', J.Acoustic.Soc.Am.**104**(1), pp. 399-409, **1998**

Churcher, B.G. and King, A.J. 'The performance of noise meters in terms of the primary standard', J.Inst.Electr.Eng.81, pp. 57-99, **1937**

Cohen, M.F. 'Detection threshold microstructure and its effects on temporal intergration data', J.Acoustic.Soc.Am.**71**, pp. 405-409, **1982**

Dai, H. 'On measuring psychometric functions: A comparison of the constant-stimulus method and adaptive up-down method', J.Acoust.Soc.Am.-**98**(6), pp. 3135-3139, **1995**

DIN/IEC 651 'Schallpegelmesser', **1881**

Evan, M.R. and Doucet, J.R. 'Is loudness simply proportional to the auditory nerve spike count', J.Acoustic.Soc.Am.**101**(5), pp. 2735-2739, **1997**

Fastl, H. und Zwicker, E. 'Lautstärkepegel bei 400 Hz-Psychoakustische Messung und Berechnung nach ISO 532B', in: Fortschritte der Akustik - DAGA 1987, pp. 189-192, Bad Honnef: DPG-GmbH, **1987**

Fastl, H., Jaroszewski, A., Schorer and E., Zwicker, E. 'Equal Loudness Contours between 100 and 1000 Hz for 30,40 and 70 phon', Acustica **70** pp. 197-201, **1990**

Florentine, M., Buss, S. and Poulsen, T. 'Temporal integration of loudness as a function of level', J.Acoust.Soc.Am. **99**, pp.1633-1644, **1996**

Florentine, M., Buss, S. and Mason, C.R. 'Level discrimination as a function of level for tones from 0.25 to 16 kHz', J.Acoust.Soc.Am.**81**, pp. 1528-1541, **1987**

Fletcher, H. and Munson, W.A. 'Loudness, Its Definition, Measurement and Calibration', J.Acoust.Soc.Am.**5**, pp. 82-108, **1933**

Gabriel, B. et al. 'Context Effects in measurements of equal-loudness level contours', in: Proceedings Inter-Noise 1995, pp. 847-850, **1995**

Gabriel, B. 'Equal-loudness Level Contours: Procedures, Factors and Models', Ph.D. thesis, Universität Oldenburg, **1996**

Gabriel, B., Kollmeier, B.and Mellert, V. 'Influence of individual listener, measurement room and choice of test tone levels on the shape of equal-loudness level contours', Acustica united with acta acustica Vol.**83**, pp. 670-683, **1997**

Garner, W.R. 'An informational analysis of absolute judgements of loudness', Journal of Experimental Psychology 46, **1953**

Green, D.M., Richards, V.M.and Forrest, T.M. 'Stimulus step size and heterogeneous stimulus conditions in adaptive psychophysics', J.Acoust.Soc.-Am.**86**(2), pp. 629-636, **1989**

Gustafsson, B. 'The loudness of transient sounds as a function of some physical parameters', Journal of Sound and Vibration, **37**(3), pp. 389-398, **1997**

Heldmann K. 'Wahrnehmung gehöhrgerechter Analyse und Merkmalsextraktion technischer Schalle', VDI-Verlag 1994, Fortschritte-Berichte, Reihe 17, Nr.109, **1994**

Hellbrück, J. 'Struckturelle Veränderungen des Hörfeldes in Abhängigkeit vom Lebensalter', Z. Gerontol. **21**, pp. 146-149, **1988**

Hellmann, R., Miśkiewicz, A. and Scharf, B. 'Loudness adaptation and excitation patterns: effects of frequency and level', J.Acoustic.Soc.Am.**101**(4), pp. 2176-2185, **1997**

Helson, H. 'Adaptation-level theory. An experimetally and systematic approach to behaviour', Harper & Row New York, **1964**

ISO 226 'Normal equal-loudness level contours', International Organisation for Standardisation, Geneva, **1987**

ISO/R 226 'Normal equal-loudness level contours for pure tones and normal threshold of hearing', International Organisation for Standardisation, Geneva, **1961**

ISO 389-7 'Reference threshold of hearing under free-field and diffuse-field listening conditions', International Organisation for Standardisation, Geneva, **1995**

ISO/TC43/WG1/N122, 'Preferred test conditions for the determination of the minimum audible field and the normal equal-loudness level contours', 3rd draft, November **1988**

Jestaedt, W. 'An adaptive procedure for subjective judgements', Perception & Psychophysics 28 (1) 85-88, **1980**

Kingsburry, B. A. 'A direct comparison of the loudness of pure tones', Phys. Rev. 29, pp. 169-177, **1972**

Kollmeier, B., Gilkey, R.H. and Sieben, U.K. 'Adaptive staircase techniques in psychophysics: A comparison of human data and mathematical method', J.Acoustic.Soc.Am.**83**, pp. 1852-1862, **1988**

Lehnhardt, E. 'Praxis der Audiometrie' (7. Auflage), Georg Thieme Verlag Stuttgart New York, **1996**

Levitt, H. 'Transformed up-down procedures in psychoacoustics', J.Acoust.-Soc.Am.**49**, pp. 467-477, **1971**

Long, G. 'The microstructure of quiet and masked thresholds', Hear.Res. 15, pp. 73-87, **1984**

Long, G. and Tubis, A. 'Investigation into the nature of the association between threshold microstructure and otoacustic emissions', Hear.Res. 36, pp. 125-136, **1988**

Lydolf, M. and Møller,H. 'New Measurements of the Threshold of Hearing and Equal Loudness Level Contours at Low Frequencies', 8th i. Meeting on Low Frequency Noise and Vibration, pp. 76-84, **1997**

Lydolf, M., Frandsen, P.C., Sørensen, M.F. and Møller, H. 'The threshold of hearing & contours of equal loudness - a study of measuring methods and normal hearing', Dissertation, Acoustics Laboratory, University Aalborg, Denmark **1999**

Mapes-Riordan, D. and Yost, W.A. 'Loudness recalibration as a function of level', J.Acoustic.Soc.Am.**106**(6), pp. 3506-3511, **1999**

Marks, L.E. 'Sensory and cognitive factors in judgements of loudness' J.-Exp.Psychol.: Human Perception and Psychophysics 5, pp.426-443, **1979**

Marks, L.E. 'Phonion: Translation and annotations concerning loudness scales and the processing of auditory intensity', in J.J. Castellan and F. Restle (Eds.), Cognitive Theory **Vol.3** ,pp. 7-31, Lawrence Erlbaum, Hillsdale, New Jersey, **1978**

Marks, L.E. 'Magnitude estimation and sensory matching', Percept.Psychophys.**43** pp.,511-525, **1988**

Marks, L.E. and Warner, E. 'Slippery context effect and critical band', J.-Exp.Psychol.: Human Perception and Psychophysics **17**, pp. 986-996, **1991**

Marks, L.E. 'Recalibrating the auditory system: The perception of loudness', J.Exp.Psychol.: Human Perception and Psychophysics **20**, pp. 382-396, **1994**

Marks, L.E. 'Recalibrating the perception of loudness: Interaural transfer', J.Acoust.Soc.Am.**100**(1), pp. 473-480, **1996**

Møller, H. and Andresen, J. 'Loudness of pure tones at Low and Infrasonic Frequencies', in: J. of Low Frequency Noise and Vibration, Vol.**3**, No.2, pp. 78-87, **1984**

Mauermann, M., Long, G. and Kollmeier, B. 'Vergleich der Feinstruktur von Isophone, Ruhehörschwelle und otoakustischen Emsissionen', in: Fortschritte der Akustik - DAGA 2000, Oldenburg, DEGA e.V., in print, **2000**

Müller,F. & Fichtl,E. 'The measurement of equal loudness contours using a direct scaling procedure and validation by equal time contours', in: Proceedings Internoise 1994, pp. 1069-1074, **1994**

Nielsen, M.K.E. and Poulsen, T. 'Hearing threshold and equal loudness level contours of 1/3- octave noise bands in free- and diffuse-field', Acustica 80, pp. 306-310, **1994**

Patterson, R.D. 'Auditory filter shape', J.Acoustic.Soc.Am.**55**, pp. 802-809, **1974**

Poulton, E.C. 'Bias in quantifying judgements', Lawrence Erlbaum, Hillscale, New Jersey, USA **1989**

Reckhardt, C. and Mellert, V. 'Modellierung von Rangeeffekten bei der Bestimmung von Isophonen', in: Fortschritte der Akustik - DAGA 1997, Oldenburg, DEGA e.V.,pp. 498 -499, **1997**

Reckhardt, C., Mellert, V. and Kollmeier, B. 'Quantifizierung der Einflussgrößen bei der Bestimmung von Isophonen', in: Fortschritte der Akustik - DAGA 1998, Oldenburg, DEGA e.V., pp. 450-451, **1998**

Reckhardt, C., Mellert, V. and Kollmmeier, B. 'Factors influencing equal-loudness level contours', Psychophysics, Physiology and Models of Hearing, T.Dau, V.Hohmann, B.Kollmeier (Editors) Word Scientific, isbn 981-02-3741-3, pp. 117-120, **1999**

Reckhardt, C., Mellert, V., Verhey, J.L. and Kollmeier, B. 'Minimizing factors influencing the results of equal-loudness level contours'. Contributions to Psychological Acoustics. Results of the 8th Oldenburg Symposium on Psychological Acoustics, Edited by August Schick, Markus Meis and Carsten Reckhardt. Oldenburg 2000, pp.129-136, ISBN 3-8142-0697-5, **2000**

Richter, U. 'Vergleich von Freifeld-Frequenzgang und Bezugs-Hörschwellenpegel von Audiometer-Kopfhörern', in: Fortschritte der Akustik - DAGA 1994, pp. 1049-1052, Bad Honnef: DPG-GmbH, **1994**

Robinson, D.W. and Dadson, R.S. 'A re-determination of equal-loudness relations for pure tones', Brit.J.Appl.Phys.**7**, pp. 166-181, **1956**

Ross, S. 'On the relation between the Acoustic reflex and Loudness', J.Acoustic.Soc.Am.**43**(4), pp. 768-779, **1968**

Schechter, M.A., Fausti, S.A., Rappaport, B.Z.and Frey, R.H. 'Age categorization of high-frequency auditory threshold data', J.Acoustic.Soc.Am.**79**, pp. 767-771, **1986**

Schloth, E. 'Relation between spectral composition of spontaneous otoacoustic emissions and fine-structure of threshold in quiet', Acustica 53, pp. 250-256, **1983**

Sone, T., Suzuki, Y., Takeshima, H., Kumagai, M., Fujimori, T., Kado, H. and Miura, H. 'New data on equal-loudness level contours for puretones and threshold of hearing in free field', in: Proceedings Inter-Noise 1990, pp. 1155-1158, **1990**

Stemplinger, I. und Gottschling, G. 'Auswirkungen der Bündelung von Verkehrswegen auf die Beurteilung der Globalen Lautheit', in: Fortschritte der Akustik - DAGA 1997, Oldenburg, DEGA e.V., **1997**

Suzuki, S., Suzuki, Y., Kono, S., Sone, T., Kumagai, M., Miura, H. and Kado, H. 'Equal-loudness level contours for pure tones under free field listening condition (I) - Some data and considerations of experimental conditions', J.Acoust.Soc.Japan(E)10, pp. 329-338, **1989**

Takeshima, H., Suzuki, Y., Kumagai, M., Sone, T., Kono, S., Fujimori, T. and Miura, H. 'Threshold of hearing for pure tone under free field conditions', J.Acoust.Soc.Japan(E)15(3), pp. 159-169, **1994**

Verhey, J.L. und Kollmeier, B. 'Messungen zur zeitabhängigen Lautheitssummation', in: Fortschritte der Akustik - DAGA 1998, Oldenburg, DEGA e.V., **1998**

Verhey, J.L. 'Psychoacoustics of spectro-temporal effects in masking and loudness perception', Dissertation, Bibliotheks- und Informationssystem der Universität Oldenburg, **1999**

Watanabe, T. and Möller, H. 'Low Frequency Hearing Thresholds in Pressure Field and in Free Field', J. of Low Frequency Noise & Vibration, Vol. **9** No.3, **1990**

Whittle, L.S, Collins, S.J. and Robinson, D.W. 'The audibility of low frequency sounds', J.Sound Vib., Vol.**21**, pp. 431-448, **1972**

Veit, I. 'Technische Akustik', Vogel Verlag, **1996**.

Yeowart, N.S.and Evans, M.J. 'Thresholds of audibility for very low--frequency pure tones', J.Acoust.Soc.Am.**55**, pp. 814-818, **1974**

Yeowart, N.S., Bryan, M.E. and Tempest, W. 'The Monaural M.A.P. Threshold of Hearing at Frequencies from 1.5 to 100 c/s', J.Sound Vib., Vol.**6**, No.3, pp. 335-342, **1967**

Zwicker, E. and Fastl, H. 'Psychoacoustics', Berlin: Springer-Verlag, **1990**

Zwicker, E. and Fastl, H. 'Psychoacoustics - Facts and Models', Berlin: Springer-Verlag, **1999**

Zwicker, E. und Feldkeller, R. 'Über die Lautstärke von gleichförmigen Geräuschen', Acustica 5, pp. 303-316, **1955**

Zwicker, E. 'Psychoakustik', Berlin: Springer-Verlag, **1982**

Erklärung

Hiermit erkläre ich, daß ich die vorliegende Arbeit selbständig verfaßt habe und nur die angegebenen Quellen und Hilfsmittel verwendet habe.

Oldenburg, den 24. August 2000

Carsten Reckhardt

Lebenslauf

Am 14.5.1969 wurde ich, Carsten Reckhardt, in Wilhelmshaven als zweites Kind von Karin Reckhardt geb. Achenbach und Karl-Jürgen Reckhardt geboren.
Von 1975 bis 1979 besuchte ich die Grundschule Jungfernbusch in Heidmühle, anschließend von 1979 bis 1981 die Orientierungsstufen in Schortens. 1981 wechselte ich auf das Mariengymnasium Jever, das ich 1988 mit dem Abitur abschloss. Im Wintersemester 1988 begann ich mit dem Studium der Physik (Diplom) an der Universität Oldenburg. 1991 habe ich die Diplom-Vorprüfung abgelegt. In den Jahren 1992 bis 1995 war ich als wissenschaftliche Hilfskraft in verschiedenen Projekten angestellt. Unter anderem habe ich bei der Erstellung von Schallgutachten und psychoakustischen Untersuchungen im itap (Instituit für technische und angewandte Physik) and in der AG Akustik mitgewirkt. Die Diplomprüfung habe ich im Mai 1995 abgelegt. Von August 1995 bis 1998 hatte ich ein Doktorandenstipendium im Graduiertenkolleg 'Psychoakustik', in dem ich meine Dissertation mit dem Titel 'Factors influencing equal-loudness level contours' begann. Von September 1998 bis April 2000 war ich Wissenschaftlicher Mitarbeiter in der Arbeitsgruppe Akustik. Anschließend wechselte ich in das Institut zur Erforschung von Mensch-Umwelt-Beziehungen unter der Leitung von Prof. Schick (von Mai 2000 bis August 2000) und arbeitete dort im Projekt 'INVITE - Zugang zum Internet für Blinde'. Seit dem 1. September 2000 bin ich als Projektleiter bei der Firma InfoOffice AG angestellt.