Preface

Whenever psychoacousticians communicate with each other, the terms "loudness", "masking", and, of course, "spectral and temporal effects" come into play. The reason for this is not only that well-known great names in auditory research (such as, e.g., von Békésy, Stevens and Zwicker) have used and defined these terms extensively. The "real" reason is in the ear itself: To a first approximation, our peripheral hearing system acts as a time-dependent spectral analyser that transforms acoustic signals with a given sound pressure level into internal auditory representations with a certain loudness. These patterns have the ability to mask out certain components of the incoming signal - an effect which is nowadays well-known and exploited in low-bit-rate storage and transmission of music (such as, e.g., the minidisk or other digital radio broadcasts). But why should a young researcher like Jesko Verhey write a complete doctoral thesis on masking, loudness perception, and spectral and temporal effects, if all these effects are already well-known and provide no special thrill to any potential reader?

Again, the answer is in the hearing system itself: Many aspects of the performance of our ear with every-day signals that are characterized by a complex spectro-temporal structure are not yet fully understood. Hence, the topic of the current thesis is to measure these effects quantitatively with psychoacoustical methods and to develop comprehensive models that try to describe as many of these effects as possible with a minimum set of assumptions. After reading this thesis you might come to the same conclusion as the members of our Graduiertenkolleg (graduate school) Psychoacoustics": Jesko Verhey has done an excellent job in obtaining new and unexpected insights into this classical area of psychoacoustics!

The first part of the thesis (Chapters 2 and 3) deals with loudness summation across frequency as a function of signal duration. Here, the term loudness summation means that the perceived loudness of a broad-band signal is much higher than for a narrow-band signal with the same energy. Since this is a consequence of the compressive nonlinearity in our hearing system, any time dependence of this loudness summation across frequency should reveal temporal properties of this compression in the auditory system - an intriguing problem that Jesko Verhey tackles with state-of-the-art psychophysical measurement methods and with an enormous amount of measurement time by his normal-hearing subjects - please read yourself!

The second part of the thesis is concerned with a more quantitative description of spectro-temporal processing using a model of the "effective" signal processing in the auditory system that has been developed before in our working group, primarily by Torsten Dau. With this model, the temporal processing within each individual frequency band is described by a modulation filterbank, i.e., a structure that separates different amplitude modulation rates (or fluctuations of the envelope, respectively) into different auditory channels. Using such a model, Jesko Verhey can solve an old paradoxon from psychoacoustics, namely the asymmetry in masking between a tone and a (narrow-band) noise: While a tone is masked rather easily by a low-level masking noise, a much higher masking level has to be applied, if a noise is masked by a tone. The explanation comes from the auditory system's capabilities of temporal analysis within each spectral band. As a consequence, more elaborate coding schemes for low-bit-rate audio transmission may become realistic - please read yourself!

Finally, the same model is employed to characterize another paradoxon of spectro-temporal processing, the so-called comodulation masking release (CMR): This refers to the ear's capability of utilizing a similar modulation in separate frequency bands (comodulation) to improve the detection of a target signal masked by this (comodulated) masker. Although this effect seems to play a major role in every-day auditory processing and auditory scene analysis, it has not been fully understood yet. Jesko Verhey can now show that only a limited part of this effect is due to an across-channel process (i.e., a process that compares different frequency channels in the temporal domain). Rather, a major part of the effect can already be explained by an appropriate temporal analysis of within-channel cues. This clearly sheds new light on the role of across-frequency comparison in the auditory system and spectral versus temporal processing. Moreover, it opens the possibility to quantitatively understand auditory perception in more complex situations with ecologically relevant signals - please read yourself!

Taken together, the current work is a milestone in the psychoacoustics of spectro-temporal effects and its quantitative modelling. It definitely will influence different areas of auditory communication such as, e.g., transmission and storage of music signals, telecommunication, and "intelligent" hearing aids. I hope that the reader will receive at least a bit of the interdisciplinary spirit of the Graduate School "Psychoacoustics" in which Jesko was involved during his thesis work: Rather than concentrating on a single effect, the approach is to take a more holistic view and to quantitatively model as many effects as possible with a single model while still observing possible technical applications. In my opinion, the current thesis is an excellent example for this approach. Admittedly, it is very satisfying and enjoyable to work with Ph.D. candidates from this interdisciplinary graduate school and, again, Jesko Verhey is an excellent representative for them. I hope that from the text and the scope of the work the reader might get an impression of how enjoyable it was to work with Jesko - please read yourself!

Oldenburg, March 1999

Birger Kollmeier

Chapter 1

General Introduction

Hearing is an important sense of the human beings to acquire information about the world in which they live. It plays a fundamental role for the communication between humans. In psychoacoustics, the relation between acoustical stimuli and the hearing sensations is investigated in order to derive a functional relationship between the physical properties of the sound and the sensations. The psychoacoustical experiments can be divided in at least two basic classes of experiments: In the first class of experiments, psychoacousticians try to directly measure the functional relationship between the physical parameters of the sound and the sensations. As an example, the subjects task is to judge the loudness of a stimulus which is primaly related to the physical magnitude intensity. In the second class of experiments, the resolution of the auditory system is investigated. In this class of experiments, detection thresholds for a certain change of the physical properties of the signal may be measured. The relation between the results of the two classes of experiments is often unclear. For example, the relation between the loudness, as determined from magnitude scaling experiments, and the data derived from intensity discrimination experiments is still subject to investigations.

The basic physical properties of sound are, apart from the intensity, its spectral content and its time structure. It is known since more than a century (Ohm, 1843, v. Helmholtz, 1863) that already the peripheral auditory system realizes a spectral decomposition, similar to a fourier analysis. Bekesy (1942, 1943, 1946, 1949a+b) showed that different frequencies excite different places on the basilar membrane in the cochlea. The frequency-place transformation can be effectively modelled by a bank of overlapping filters. Fletcher (1940) was the first to interpret his detection data assuming a critical-band filterbank. He measured the detection threshold for sinusoidal signals in the presence of a bandpass-noise masker as a function of the bandwidth of the masker. He concluded from his results that (i) the peripheral auditory system consists of a bank of overlapping critical-band filters, and (ii) the thresholds are determined by the overall energy of the masker in the peripheral filter centered at the signal frequency. Based on these assumptions, Fletcher proposed a model known as the power-spectrum model. Later on, it was shown by several authors that the power spectrum model can predict the results of several detection experiments.

Also for loudness perception, the frequency-place transformation seems to play an important role. It has been shown by several authors that, keeping the sound pressure level of a complex stimulus constant, the loudness level of the stimulus can significantly increase when the bandwidth exceeds the critical bandwidth (e.g., Fletcher and Munson, 1933, Scharf, 1959, Zwicker et al., 1957). This effect has been called loudness summation. In general, loudness models account for this summation effect by assuming three processing stages: A bank of (critical-band) filters, a compression at the output of each critical band, and finally a summation across frequency bands (Zwicker and Scharf, 1965).

Zwicker's loudness model and the power-spectrum model work well for stationary signals. However, most of the surrounding sounds are of nonstationary nature, i.e, there intensity varies over time. A prominent example is speech. Therefore, it can be expected that the auditory system is also using temporal information of the sound.

In dynamic loudness models, it is generally assumed that the growth and decay of loudness can be described by certain time constants. However, there is considerable confusion in the literature as to the exact numerical value of the time constant. For the building up of loudness, time constants in the range from 25 ms (Niese, 1959) to 1 s (Sone et al., 1986) were proposed. As in the case of the build-up time constant, many different values were proposed for the decay-time constant, in the range from 250 ms (Kumagai et al., 1984) to 5 s (Ogura et al., 1993). This indicates that further investigation may be necessary to better understand the temporal processing of loudness. For example, models of dynamic loudness generally assume that spectral integration of loudness (loudness summation) is independent of duration. However, the data of the few publications concerning temporal aspects of loudness summation are contradictory. Whereas Port (1962, 1963b) and Zwicker (1965) found the same amount of loudness summation for short and long signals, Boone (1972) measured a duration-dependent loudness-summation effect.

The overview given so far points out the interaction between the processing of intensity, spectral content and temporal effects in auditory psychophysics, which is still not fully understood.

The current work therefore aims at clarifying parts of this intricate interaction: The first part of the thesis (Chapters 2 and 3) investigates the role of temporal processing in loudness experiments by measuring the loudness-summation effect at moderate levels as a function of signal duration. Since loudness measurements are often, unintentionally, influenced by methodological factors (such as, e.g., the level range of the signals presented in the experiments)(e.g., Gabriel, 1996), their influence on the measurement of spectral integration of loudness with an adaptive procedure is quantified separately in Chapter 2. In Chapter 3, loudness is measured as a function of signal bandwidth and duration with an optimized procedure, which is based on the findings of Chapter 2.

The second part (Chapter 4 and 5) investigates the role of temporal processing in detection experiments. Experiments from the literature are replicated and additional experiments are performed and compared to simulated data. Simulations are performed with the model of the "effective" signal processing in the auditory system presented by Dau et al. (1997a+b), which was originally developed to account for modulation detection and masking data. In Chapter 4, measurements and simulations of spectral masking are presented. First, the ability of the model to account for the frequency selectivity of the auditory system is tested. Then, the role of temporal effects in "asymmetry of masking" conditions is investigated, where it is observed that thresholds for test signals with bandwidths larger than the masker bandwidth are much lower than those for the reversed condition, although all bandwidths are smaller than the critical bandwidth. The power-spectrum model would predict the same threshold for both conditions.

Another phenomenon, which cannot be predicted by the power spectrum model, is the difference in the masking properties of modulated and unmodulated maskers. It was shown in the literature, that the detection threshold for a test signal presented in a noise masker can be markedly lower when the masker is modulated compared to the condition where the masker is unmodulated. The effect was called "comodulation masking release" (CMR), since the threshold difference between the two conditions is large for bandwidths larger than the critical bandwidth and it was suggested that the observer may compare the correlated envelope information across frequency bands in the modulated condition (across-frequency process). However, some authors (e.g., Schoonefeldt and Moore, 1989a) argued that part of the CMR-effect may be explained by within-channel cues, where only the output of the peripheral channel centered at the signal frequency is processed. In Chapter 5, the role of within-channel cues which has has often been discussed is quantified in one class of CMR-experiments (characterized by a single bandpass noise centered at the signal frequency). Simulations obtained with a single-channel modulation-filterbank model are performed and compared to experimental results.

It will be shown in this thesis that spectro-temporal effects in masking experiments can be modeled with comparatively few assumptions, whereas spectro-temporal effects in loudness perception seem to be more complicate. In the latter case, the findings of the current thesis indicate that a new modeling approach is needed to predict the loudness of temporally varying sounds. Ideas concerning the important aspects of such a new model will be proposed in this thesis.

Chapter 2

Influence of methodological factors on estimates of spectral loudness summation[1]

ABSTRACT

The intention of the present study is to quantify and minimize the influence of methodological factors on the measurement of spectral integration of loudness (loudness summation). Equal-loudness-level differences for bandpass-noise signals spectrally centered at 2 kHz are measured as a function of the bandwidth in the range from 200 to 6400 Hz using an adaptive 2-interval 2-alternative forced-choice 1-up 1-down procedure. The influence of the following methodological factors are investigated: (i) the level difference between test and reference signal at the beginning of an adaptive track (starting level), (ii) the procedure using interleaved tracks in comparison to a simple adaptive procedure (non-interleaved), (iii) the reference bandwidth. When non-interleaved tracks were used for the measurement of loudness summation for short signals (10 ms), no effect of the starting level is observed, whereas, for long signals (1000 ms), the starting level has a large effect on measured equal-loudness levels. Using interleaved tracks, the starting level does not have any influence on the equal loudness levels for long signals. Furthermore, when interleaved tracks are used, the same amount of loudness-summation is obtained for the two different reference bandwidths tested (400 and 3200 Hz). The results of the present study indicate that an adaptive procedure with non-interleaved tracks is not appropriate to measure temporal aspects of loudness summation, because the effect of the starting level on equal loudness levels can not be separated from the effect of the physical parameter duration. Instead, interleaved tracks should be used to avoid methodological factors influencing the results.

[1] This chapter is a slightly modified version of Verhey, J.L., and Kollmeier, B. (1998) “Influence of methodological factors on estimates of spectral loudness summation”, submitted to JASA.

2.1 INTRODUCTION

Loudness is a fundamental psychoacoustical attribute of a sound. It is defined as the subjective intensity of the sound, independent of any meaning the sound might have (Kryter, 1985). Loudness is primarily related to the sound intensity. In addition, loudness depends on other physical properties of the signal such as duration and spectral content of the signal. One way to get information about loudness perception is to measure the level of signals with different physical properties producing the same loudness sensation (equal-loudness levels). Equal-loudness levels can be measured directly in loudness-comparison experiments or indirectly by comparing the results of loudness scaling of the stimuli with different physical properties.

Loudness measurements are often, unintentionally, influenced by methodological factors (such as, e.g., the level range of the signals presented in the experiments). Methodological factors can affect both loudness-scaling and loudness-comparison data. Schneider and Parker (1990) used a procedure where the subjects heard two pairs of tones and responded by indicating which pair differed more in loudness. They found that changing the level range in the experiment can change the perceived loudness by the same amount as reported by Marks (1988) and Marks and Warner (1991) for magnitude estimation. The level range of the signal can also have a large effect on the equal loudness levels measured with a constant stimuli method (Gabriel, 1996, Gabriel et al., 1997, Marks, 1994). A particularly large effect was reported by Marks (1994). He found that the equal loudness level between a 500-Hz and a 2500-Hz signal can change by as much as 24 dB when different level ranges of the signals were used. This is even larger than the level difference expected from the ISO 226. Marks (1994) and Gabriel et al. (1997) reported that the mean of the level range influences the results. Gabriel (1996) concluded from her results that the highest test-tone level presented within a run affect the resulting point of subjective equality more than lower test-tone levels. In agreement with Marks (1994), she found an increasing effect of the level range with increasing frequency difference.

This finding indicates that a change in physical properties of the reference could change the results in loudness-comparison experiments. In fact, it was shown in the literature that the choice of the reference signal can markedly influence the results. Several authors reported an influence of the reference duration on temporal integration of loudness. For example, Takeshima et al. (1988) found that the level difference between equally loud signals with a duration of 0.05 and 10 sec is 4 dB when a reference signal with a duration of 1 sec was used whereas it is 9 dB when the reference duration was 0.2 msec. Reichardt and Niese (1970) and Reichardt (1970) tried to optimize the properties of the reference signal in measurements of temporal integration of loudness. They investigated how the difference between the duration of the reference and the test signal influence the distribution of the individual equal-loudness levels from 50 subjects. They found that the statistical spread increases with increasing difference between test-signal and reference-signal duration. They concluded that for large differences in duration it is almost impossible to establish a mean value. Therefore, according to Reichardt, only test-signal reference pairs with small differences in duration should be used. For spectral integration of loudness, Boone (1973) reported about a similar problem. He found an unacceptable spread among the test subjects when large test-signal bandwidths and a sinusoidal reference signal were used.

In general, methodological factors affect the results considerably, when the difference between the physical parameters of the signals (such as duration or spectral content) is large, whereas almost no influence of methodological factors is observed when the signals have the same physical properties (Reichardt and Niese, 1970, Reichardt, 1970, Marks, 1988, Marks, 1994, Gabriel, 1996). It is desirable to separate the influence of the methodological factor from the influence of the variation of the physical parameters of the sound.

As mentioned above, several studies investigated the influence of methodological factors in loudness measurements on *temporal* integration or loudness matching experiments with sinusoidal signals. However, only a few publications reported about methodological factors in measurements on *spectral* integration of loudness (loudness summation). For example, Marks (1978) measured spectral integration of loudness for two tones, one at 2000 Hz and the other at 5000 Hz, using magnitude estimation. He found that his data contradict the findings of other authors, that the total loudness of the two-tone complex is simply the sum of the loudness attributed to each of the two tones, when the frequency difference of the two tones is large. The total loudness was less than the sum of the loudness of the two sounds. To explain his results, he proposed that the magnitude estimates of the total loudness were biased. To obtain additivity, he proposed to raise the magnitude estimates of the total loudness to a power of 1.33. Such a transformation is consistent with a two-stage model of magnitude estimation in which sensation is a power function of intensity and numerical estimates are a power function of sensations (Rule and Curtis, 1982, Marks, 1979, 1994). Schneider (1988) argued that it is difficult to test for additivity of loudness from different spectral regions when one believes that the results are biased. He concluded, that a different procedure such as loudness comparison should be used. Hübner and Ellermeier (1993) proposed a two-interval, adaptive forced-choice procedure with interleaved tracks to test additivity of loudness. A similar procedure was proposed by Florentine et al. (1996) to measure temporal integration of loudness. They showed, that at least for temporal integration the effect of some methodological factors (such as the physical properties of the reference) are minimized. Unfortunately, Hübner and Ellermeier did not test if the their results on spectral integration of loudness were influenced by methodological factors (such as e.g., starting level, reference).

In the present study, the influence of the several methodological factors on loudness-summation experiments are investigated using an adaptive procedure: First, the influence of the level difference between test and reference signal at the beginning of an adaptive track (starting level) on the results is examined. Second, the effect of the procedure using interleaved tracks on the results in comparison to a simple adaptive procedure (non-interleaved) is considered. Finally, the influence of the reference bandwidth is investigated.

2.2 METHODS

2.2.1 Procedure

Stimuli with different bandwidths were matched in loudness to a reference signal with a fixed bandwidth using an adaptive 2–interval, 2–alternative forced choice procedure. In each trial the listeners heard two sounds, the reference and the test signal, which were separated by an 500 ms silent interval. Test and reference signal were presented in random order and with equal a priori probability. The listeners indicated which signal was louder by pressing the corresponding key on a keyboard. The reference level was fixed and the level of the test signal was varied according to a simple 1–up 1–down procedure, which converges at the 50 % point of the psychometric function (Levitt, 1971). If the listener indicated that the test signal was the louder one, its level was reduced in the next trial, otherwise it was increased. At the beginning, the step size was 8 dB. It was divided by two when the listeners indicated in one trial that the test signal was louder and in the previous trial that the reference signal was louder. A minimum step size of 2 dB was maintained for the final 4 reversal of the track. The equal-loudness level for each track was determined by calculating the median level during these 4 reversals. The level difference between test and reference signal at the beginning of an adaptive track (starting level) was either 20 dB above, 20 dB below, or at the reference level (0 dB difference). Four tracks were run for each listener, pair of stimuli and starting level.

In the first experiment, non-interleaved tracks were employed, i.e. the adaptive tracks for the different stimulus pairs were performed subsequently. Hence, the adaptive track for one stimulus pair has to be finished before the adaptive track for the next stimulus pair starts. The order of stimulus pairs was presented randomly. The starting level was the same for all tracks in a series of trials.

In the second experiment, several interleaved adaptive tracks were run concurrently in a series of trials to simultaneously obtain loudness matches for all stimulus pairs tested. Such a procedure was proposed in the literature to reduce biases that occur when only one stimulus pair was matched in loudness (Florentine et al., 1996, Hübner and Ellermeier, 1993). On each trial, the track was chosen randomly from all possible tracks, i.e., from all tracks that had not yet been terminated. To ensure that the interleaved tracks converge at roughly the same time, the random choice of tracks is further restricted by the following rule: If none of the tracks is terminated, the number of trials for all different tracks has to be the same before the next trial for all tracks is presented in random order to the listener. If one track is terminated, the rule is applied to the choice of trials from the remaining unterminated tracks. In a series of trials, the tracks were equally distributed across the three starting levels mentioned above. All starting levels had the same frequency of occurrence in a series of trials.

In general, the procedure is comparable to the procedure proposed by Florentine et al. (1996). However, there are some details which differ considerably. First, the additional rule for choosing the next trial was not incorporated in the procedure proposed by Florentine et al. (1996). Second, in Florentine et al. (1996), different reference signals were used in a series of trials. In the present study, the reference signal was the same within each series of trials and its level was fixed. Another difference is the choice of the starting level. Whereas in the present paper the tracks started at 20, 0, or -20 dB relative to the reference level, Florentine et al. used a starting level of 10 dB above the expected value. The latter two differences originate from a different motivation of the two studies. Whereas in the present study, the influences of the starting level and the reference bandwidth are investigated explicitly, the major goal of Florentine et al. was to minimize (and not to quantify) possible bias effects. Therefore, in the present study, an attempt was made to keep the procedure as simple as possible.

2.2.2 Apparatus and stimuli

The stimuli were generated digitally with a sampling rate of 32 kHz. Stimulus generation and presentation were controlled by a silicon graphics workstation (INDY), which also sampled the listener's responses and controlled the procedure. The software package SI, which was developed at the University of Göttingen, was used for signal generation and controlled the experiments. The stimuli were D/A converted (16 bits), and then preamplified and lowpass-filtered at 16000 Hz with a computer controlled audiometric amplifier developed in a joint research project on speech audiometry (Kollmeier, 1996). The subjects sat in a double-walled, sound-attenuating booth. The stimuli were presented diotically via a Sennheizer HD 25 headphone without free-field equalization.

Bandpass-noise signals with a flat spectrum were used. To generate the signals, a broadband noise was digitally filtered by setting the magnitude of the Fourier coefficients to zero outside the desired passband. The bandpass noise was geometrically centered at 2 kHz. The bandwidths of the test signals were 200, 400, 800, 1600, 3200, and 6400 Hz. In the first experiment, the reference bandwidth was 400 Hz and the level of the reference signal was fixed at 55 dB. In the second experiment, in addition to the reference used in the first experiment, a second reference with a bandwidth of 3200 Hz and a level of 45 dB was tested. Test signals and reference signal had the same duration. It was either 10 or 1000 ms. Signals were gated with 2.5 ms raised cosine ramps.

2.2.3 Subjects

Six subjects (4 male, 2 female) participated in the experiments. They ranged in age from 25 to 28. Five of them were members of the "Graduiertenkolleg Psychoakustik" or the "Arbeitsgruppe Medizinische Physik" at the University of Oldenburg, one of them being the author JV. One subject (AR) received a payment for the participation in the experiments. All subjects had normal hearing (absolute threshold in quiet $\leq$ 15 dB HL) and no previous history of any hearing problems.

2.3 RESULTS

2.3.1 Experiment 1: Loudness summation measurements using non-interleaved tracks

Figure 2.1 shows the level differences between a test and a reference signal to obtain equal loudness (equal-loudness level-differences) as a function of the test-signal bandwidth. The test signal and the reference signal had a duration of 1000 msec. The reference bandwidth was 400 Hz. A negative level difference means that, to produce the sensation of equal loudness, the test signal must have a lower level than the reference. Parameter is the difference between test-signal level and reference-signal level at the beginning of an adaptive track (starting level). Three starting levels were tested: 20 dB (circles), 0 dB (squares), and -20 dB (triangles). Individual data are shown in the upper four and in the lower left panel. The error bars represent plus minus one standard deviation calculated across the four repetitions for each subject and condition. In the lower right panel the averaged data across subjects are shown. In this panel, the vertical bars show the amount of variability across subjects, calculated as plus minus one (interindividual) standard deviation of the means across the five subjects.

As expected, all subjects show no level difference when test and reference signal have the same bandwidth (400 Hz). Large negative level differences were obtained for large test-signal bandwidths (3200 and 6400 Hz), This finding is in accordance with the data in the literature (e.g. Zwicker et al., 1957, Port, 1963a). However, there are large individual differences in the shape of the function. Whereas for three subjects (KT, OW, and JV), level differences between test and reference signal decrease continuously with increasing test-signal bandwidths > 400 Hz, subject CR show (negative) level differences which are almost independent of test-signal bandwidth for bandwidths $\geq$ 800 Hz. For subject JA, level differences are independent of bandwidth in the range from 400 to 1600 Hz and decrease only for test-signal bandwidths > 1600 Hz. For all subjects, equal-loudness level differences are the same or slightly increase with increasing test-signal bandwidth (1 dB on average) between 200 and 400 Hz.

Obviously, there are large individual differences among subjects in the amount of the loudness summation effect. The equal loudness level difference between reference and test signal with the largest bandwidth (6400 Hz) ranges from 3 dB (subject CR, starting level 20 dB) to 17 dB (subject KT, starting level -20 dB). For all subjects, the measured loudness summation effect depends on the starting level. On average, loudness summation is 12.5 dB for a starting level of -20 dB, 10 dB for a starting level of 0 dB, and 7.5 dB for a starting level of 20 dB, i.e., loudness summation is up to 5 dB larger for a starting level of -20 dB than for 20 dB (see lower right panel of Fig. 2.1). The difference between the results obtained with a starting level of 20 dB and -20 dB is highly significant

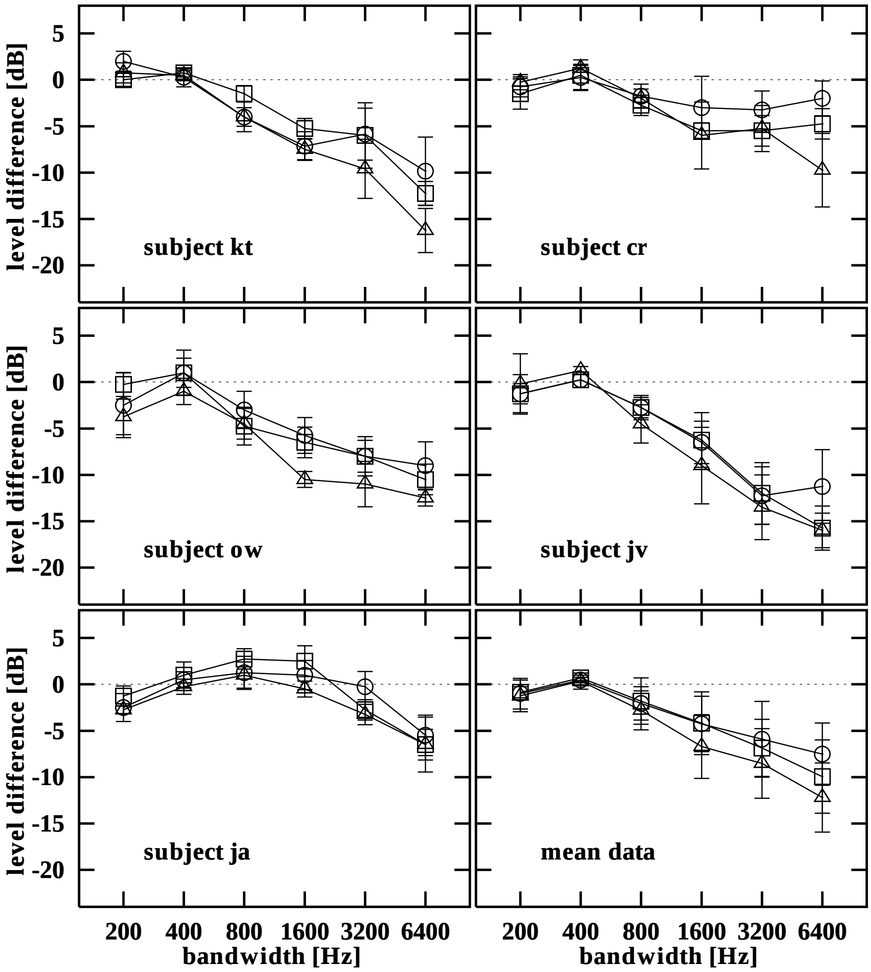

Figure 2.1: *Equal-loudness levels measured with non-interleaved tracks as function of test-signal bandwidth for a starting level of 20 dB (○), 0 dB (□), and -20 dB (△). Signal duration was 1000 ms. The reference bandwidth was 400 Hz and the level of the reference was fixed at 55 dB. The upper four panels and the lower left panel show individual data with standard deviation. The lower right panel show averaged data across subjects with interindividual standard deviations.*

for a test signal bandwidth 3200 and 6400 Hz (Student Test , $p < 0.01$).

Interindividual standard deviations increase continuously with increasing difference between reference and the test-signal bandwidth (up to 4 dB). This reflects the large individual difference in the amount of the loudness-summation effect.

The intraindividual standard deviations are smaller than 4 dB. For most subjects,

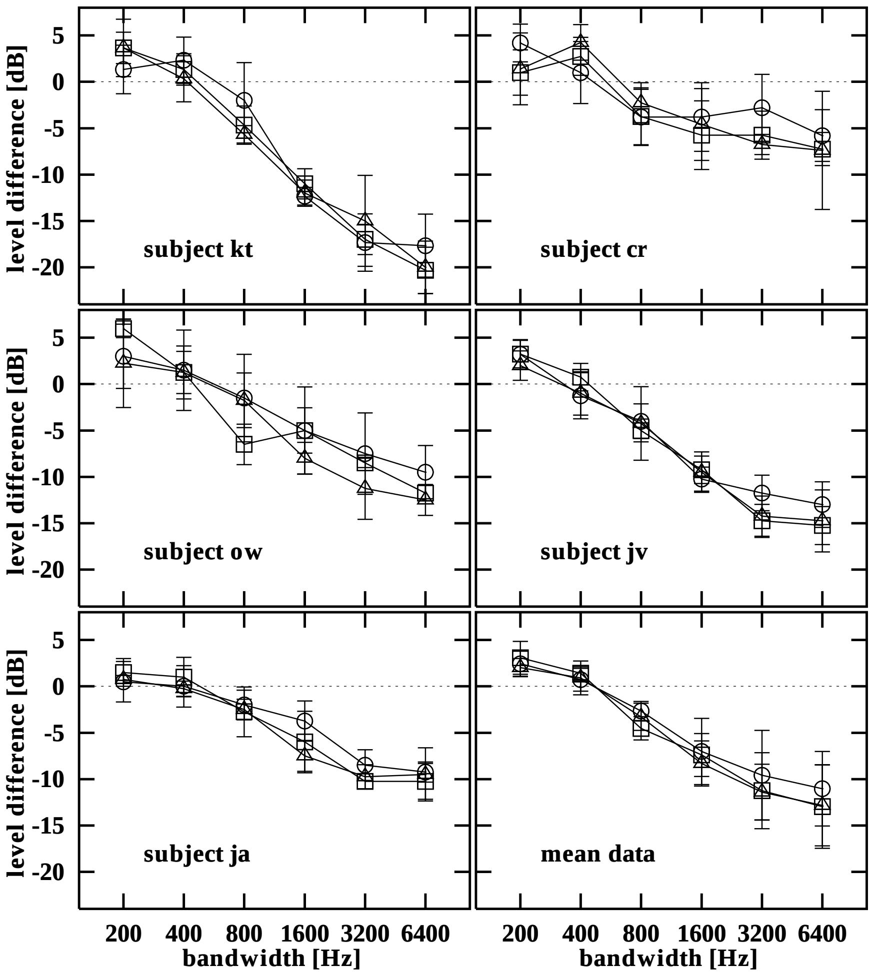

Figure 2.2: *Identical to Fig. 2.1, but for a signal duration of 10 ms.*

standard deviations are large for large test-signal bandwidths (3200 Hz, 6400 Hz) and small for small differences between reference and test-signal bandwidth. This is consistent with the subjective reports of the subjects who found it very difficult to equate the loudness of signals where the bandwidths differ considerably. However, the interindividual standard deviation does not increase monotonically as the difference between test and reference bandwidth increases as found for the averaged data. For example, subjects CR and JV show smaller standard deviations for a test-signal bandwidth of 3200 Hz than for 1600 Hz.

Figure 2.2 shows equal-loudness level differences between test and reference signals for a signal duration of 10 msec. The level differences are plotted in the same manner as in Fig. 2.1. No level difference between equally loud test and reference signals occurs when test and reference signal have the same bandwidth. Level differences decrease continuously with increasing test-signal bandwidth for bandwidths > 400 Hz. Qualitatively, this is consistent with the data for the 1000-msec long signals. However, in some conditions the data differ considerably. First, in contrast to the data obtained with long signals, positive level differences are obtained for a test-signal bandwidth of 200 Hz (2 dB on average). This is the case for all subjects and starting levels. Second, on average, the level difference between equally loud reference and test-signal with a bandwidth of 6400 Hz is larger for 10-msec than for 1000-msec long signals. This duration-dependent difference depends on the starting level. It amounts to about 3.5 dB for a starting level of 20 dB, 3 dB for a starting level of 0 dB, and only 0.5 dB for a starting level of -20 dB. Third, effect of the starting level on the loudness summation measurement is not significant ($p > 0.05$[2]). On average, the difference between the results for the different starting levels does not exceed 2 dB.

As already observed for the results with the 1000-msec test-signal (c.f. Fig 2.1), the interindividual standard deviations increase with increasing difference between reference and test-signal bandwidth by up to 4 dB. Because of the short duration of the signals, most subjects reported difficulties to perform the experiments. This is probably the reason for slightly increased intraindividual standard deviations (up to 7 dB) for some subjects (OW, CR, KT). As above, no strict monotonous relation is found between the intraindividual standard deviation and the test-signal bandwidth.

In summary, the maximum loudness summation effect (defined as the difference between maximal and minimal level) is larger for 10-msec than for 1000-msec signals. However, the temporal effect of loudness summation can not be separated from the effect of starting level.

2.3.2 Experiment 2: Loudness summation measurements using interleaved tracks

In the previous section, it was shown that loudness summation using *non-interleaved* tracks depends on the starting level for long signals, whereas it is almost independent of starting level for short signals. Therefore, for long signals, the effect of the starting level on loudness summation measurements was tested using *interleaved* tracks.

Figure 2.3 shows equal-loudness level-differences between test and reference signal as a function of the test-signal bandwidth. The data were obtained with an adaptive procedure using interleaved tracks. The signal duration was 1000 msec. The level differences are plotted in the same manner as in the previous figures.

The general shape of the functions shown in Fig. 2.3 is the same as in the previous figures. However, quantitatively they differ considerably from the data for long signals obtained with non-interleaved tracks (cf. Fig. 2.1). First, for all subjects loudness summation is larger than found in Experiment 1. The maximum increase amounts to 9 dB

[2]Except for a test-signal bandwidth of 800 Hz the starting level has no significant effect on the results with $p>0.2$.

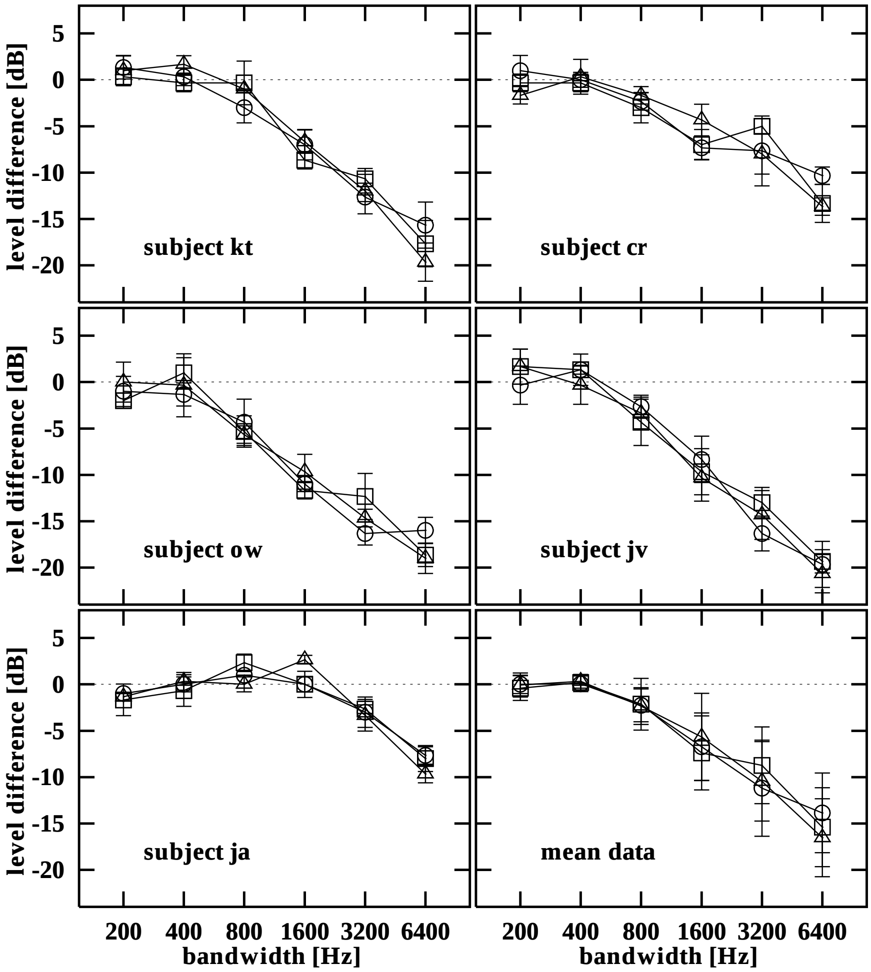

Figure 2.3: *Equal-loudness levels measured with interleaved tracks as function of test-signal bandwidth. The level differences needed to obtain equal loudness between test and reference signal are plotted in the same way as in Fig. 2.1. Signal duration was 1000 ms. The reference was the same as in the previous figures.*

(subject CR). On the average across all subjects and starting levels, the level difference between the reference and the test signal with the largest bandwidth (6400 Hz) is 15 dB when interleaved tracks were used, whereas it is only 10 dB when non-interleaved tracks were used. Note, however, that the individual differences in the amount of loudness summation effect remain: As in the previous experiments, subjects CR and JA show small loudness-summation effect and subject JV and KT show the largest loudness-summation effect. As a consequence, interindividual standard deviations are the same for both procedures (cf. lower right panel of Fig. 2.1 and Fig. 2.3). However, the variations in the individual data in Fig. 2.3 is smaller than in Fig. 2.1 (intraindividual standard deviation < 3 dB).

A second major difference is the effect of the starting level. In fact, there is no systematic variation of the measured loudness summation effect with respect to the starting level as was found for the data for long signals using non-interleaved tracks. On the average across subjects, the difference between equal-loudness level differences for different starting levels does not exceed 3 dB for a stimulus pair and is not significant ($p > 0.2$). Thus, the starting level has no influence on loudness summation measurements when interleaved tracks are used. As a consequence, only mean results averaged over the three starting levels are shown in the next figures.

Figure 2.4 shows for six listeners individual equal loudness level differences between test and reference signal as a function of test-signal bandwidth for two reference bandwidths. Again, the signal duration was 1000 msec. Except for AR (lower right panel), all subjects participated in the previous experiments. Figure 2.5 shows averaged data across the six subjects. The circles represent level differences for a (narrowband) reference signal with a bandwidth of 400 Hz at a level of 55 dB (as used in the previous experiments) whereas the squares represent level differences for a (broadband) reference signal with a bandwidth of 3200 Hz at a level of 45 dB.

As expected, there is no level difference between test and reference signal in the case, where both signals have the same bandwidth, i.e., at 400 Hz for the narrowband and at 3200 Hz for the broadband reference signal, respectively. As a consequence, the equal loudness levels for the broadband reference signal are shifted towards positive level differences. The shape of the two curves differs across subjects. Whereas for subject JA the two curves tend to converge at large test-signal bandwidths, for the other subjects the two curves are more or less parallel (Fig. 2.4). On the average across subjects, there is only a constant shift of 10 dB but no further difference in shape of the curves for the two reference signals (cf. Fig. 2.5).

The intraindividual standard deviations depicted in Fig. 2.4 are rather small. For most data points, they are less than 2 dB. No standard deviation exceeds 3 dB. In agreement with the previous results, inter-individual standard deviations increase for increasing differences between reference and test-signal bandwidth. Large standard deviations (up to 4.5 dB) are observed for narrowband test-signals in the case of a broadband reference signal and for broadband test-signals in the case of a narrowband reference signal, respectively (cf. Fig. 2.5).

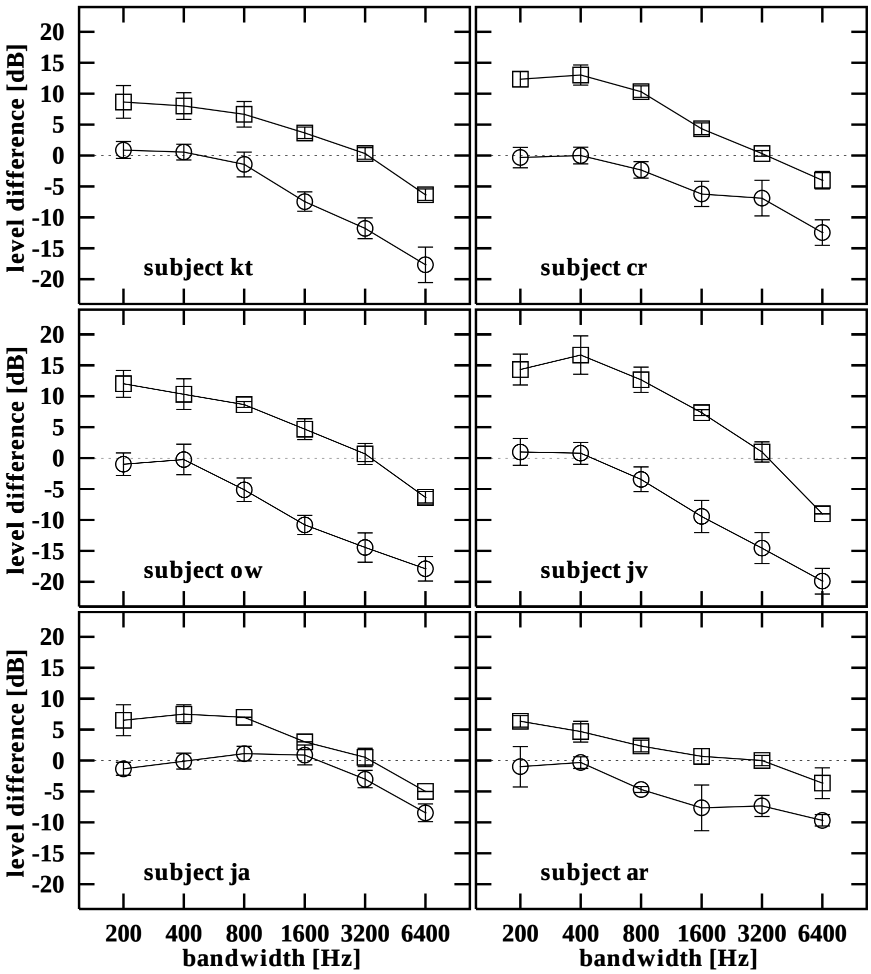

Figure 2.4: *Equal-loudness levels measured with interleaved tracks as function of test-signal bandwidth from six listeners. Signal duration was 1000 ms. Two reference bandwidth were tested: 400 Hz (○), and 3200 Hz (□). The vertical bars show plus and minus one standard deviation of the mean.*

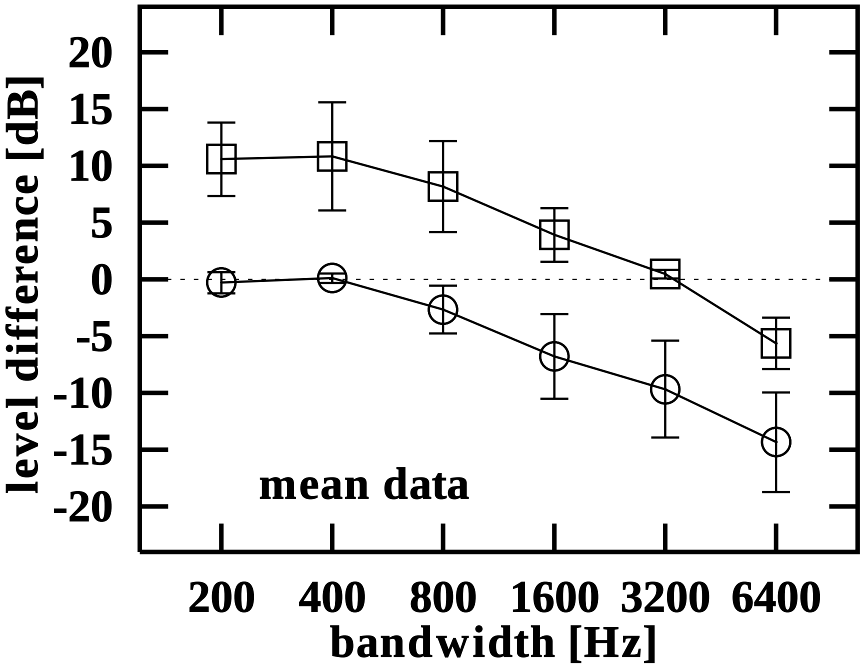

Figure 2.5: *Averaged equal-loudness levels measured with interleaved tracks as function of test-signal bandwidth for two reference bandwidths from six listeners. Symbols are the same as in Fig. 2.4. The vertical bars show plus and minus one standard deviation of the means across the six subjects.*

2.4 DISCUSSION

In the present study, the effect of the starting level was tested for different signal durations and different procedures. It was shown that the starting level can influence the results when a simple (non-interleaved) procedure is used. This influence was highly significant ($p < 0.01$) for long signals and large differences between test-signal and reference bandwidth. Compared to the data obtained with a starting level of -20 dB, the amount of loudness summation is decreased by about 5 dB when a starting level of 20 dB is used. Recently, Mellert and Reckhardt (1997) and Reckhardt et al. (1998) reported a similar effect of the starting level on equal loudness levels between tones with different frequencies measured with the same procedure. The effect of the starting level could reflect the tendency of the subjects to use responses equally often (e.g., Pollack, 1964, Stevens and Galanter, 1957). In fact, a starting level of -20 dB is near the measured equal loudness level, whereas the starting level of 20 dB is about 30 dB above the point of subjective equality, i.e., the measured equal loudness level might be increased.

Poulton (1977, 1989) suggested that range effects which are observed when the method of constant stimuli is used might be due to the response behaviour. When subjects know which stimulus is the reference stimulus, they might ignore the reference after a while, but judge each signal against the range of stimuli presented. However, a range effect as found for constant stimuli method is unlikely to be observed when an adaptive procedure is applied, because in this case, the level range is not fixed. Instead, the level range will be determined by several factors, such as the starting level, the step size, and the subjective responses. Nevertheless, we can not rule out that also in the present experiments the subject could show the tendency to ignore the reference after a while. Then, as found in loudness scaling experiments, the order of the test-signal levels within an adaptive track might influence the results (context effect). It has been shown by Hohmann et al. (1997) in categorical scaling experiments that for ascending order of levels the perceived loudness in the midlevel range is increased compared to the loudness values, obtained with a descending order of levels. In an adaptive procedure, the order of levels is not fixed, but determined by several factors of the procedure, including the response behaviour of the subject. However, when the starting level differ considerably from the equal-loudness level then such an ascending or descending order of levels may occur at the beginning of the experiment. Thus, for very high starting levels, the equal-loudness levels may be higher than those for very low starting levels. This is in accordance with the results of the present study. At the moment, both the tendency to use responses equally often and the tendency to ignore the reference combined with the context effect are possible explanations for the starting-level effect.

Interestingly, the effect of the starting level depends on the duration of the signals. For short signals, there is no significant effect of the starting level on the amount of loudness summation. At the first view, this finding seems to contradict the hypothesis that the effect of the starting level results from the tendency to use the responses equally often or that the level order for extreme starting levels influences the results. Although subjects reported that they generally found it very difficult to judge the loudness of short signals (10 msec), they also reported that the short signals are more similar to each other than the long signals (1000 msec). This might explain the reduced methodological effect for the short signals.

When interleaved tracks are used, subject reported that they were unable to follow the different adaptive tracks. Therefore, range effects may play a minor role. Because different starting levels are used for different tracks, it is unlikely that monotonically ascending or descending levels occur. In addition, an effect due to the tendency to use responses equally often will be minimized, because some of the tracks will start above the point of subjective equality and some below it. Thus, for an adaptive procedure, the use of interleaved tracks should minimize the influence of methodological factors. This was also found in the present study, where for long signals the effect of the starting level disappears when interleaved tracks are used. Note however, that randomization per se does not guarantee that the results are not influenced by methodological factors. For the method of constant stimuli, the level range of the signals still influence the results significantly, even when several test-tone frequencies were measured concurrently within a run (Gabriel, 1996).

As already mentioned in the introduction, the choice of the reference can markedly change the results. To our knowledge, there is only one publication concerning the influence of reference in loudness experiments using an adaptive procedure with interleaved tracks. Florentine et al. (1996) measured temporal integration of loudness and compared the directly measured level differences between equally loud 5- and 200-msec stimuli with those obtained indirectly by adding level differences between 5- and 30-msec stimuli and between 30- and 200-msec stimuli. They showed that using an interleaved procedure, on average, the discrepancies between direct and indirect measurements are small (≤ 3 dB) although individual data vary considerably. The finding suggests, that the influence of the reference is markedly reduced, when a procedure with interleaved tracks is used. Qualitatively, this is in agreement with the results of the present study. They show a negligible influence of the reference bandwidth on the amount of loudness summation.

Taken together, it was shown in the present study that methodological factor have no significant effect on the data when interleaved tracks were used, whereas they affect the results when a simple adaptive procedure is used.

2.5 SUMMARY AND CONCLUSION

The starting level has no significant influence on the results for short signals (10 msec). However, it influences the amount of loudness summation for long signals (1000 msec) when non-interleaved tracks are used. The effect increases (up to 5 dB) with increasing difference between the test-signal and reference bandwidth. Using interleaved tracks, the effect of the starting level for long signals disappears. No effect of the reference bandwidth on the amount of measured loudness summation was found with interleaved tracks.

The results of the present study indicate that loudness summation should be measured with an adaptive procedure using interleaved tracks. This is particularly important when temporal aspects of loudness summation are investigated. Using non-interleaved tracks, the influence of methodological factors can not be separated from the effect of the variation of physical parameters of the signals.

Chapter 3

Loudness summation as a function of duration [1]

ABSTRACT

Loudness was measured as a function of signal bandwidth and duration. The bandwidth was varied in the range from 200 to 6400 Hz and duration was varied in the range from 2.5 to 1000 ms. The signal was either a bandpass noise or a chirp train spectrally centered at 2 kHz. The level to produce equal loudness was measured with an adaptive, two-interval, two-alternative forced-choice procedure. A loudness balancing procedure was used, where the tracks for all signal pairs to be compared were interleaved. Mean results over ten normal hearing subjects showed that the loudness-summation effect for bandpass noise signals depends on signal duration. The level difference between equally loud signals with the smallest ($\Delta f = 200$ Hz) and largest ($\Delta f =$ 6400 Hz) bandwidth decreases for increasing duration in the range from 10 to 1000 ms.

Chirp trains were used to examine loudness summation for signal durations smaller than 10 ms. Mean results over six subjects showed that the loudness summation effect for chirp trains is independent of signal duration in the range from 2.5 to 40 ms. No further increase of loudness summation was found in this range of duration. However, the findings for the two types of signals are not directly comparable, because the perception of the chirp trains is different from that of the noise signals and the range of bandwidths (from 1600 to 6400 Hz) tested is rather small.

The results are discussed in terms of a duration-dependent compression within each peripheral filter.

[1] This chapter is a slightly modified version of Verhey, J.L., and Kollmeier, B. (1998) "Loudness summation as a function of duration", submitted to JASA.

3.1 INTRODUCTION

The physical magnitude intensity is related with the hearing sensation loudness. However, loudness depends not only on the intensity of the signal, but also on other physical properties of the signal such as duration and spectral content of the signal.

Many studies have investigated how loudness changes with the bandwidth of the signal. It was shown that a broadband signal sounds louder than a narrowband signal of the same physical level. Keeping the sound pressure level of a complex stimulus fixed, the loudness level of the stimulus can significantly increase when the bandwidth exceeds the critical bandwidth (e.g. Fletcher and Munson, 1933, Scharf, 1959, Zwicker et al., 1957). This effect was called loudness summation. In general, loudness models account for the effect by assuming three processing stages: A bank of (critical-band) filters, a compression within each critical band, and a summation across frequency bands (Zwicker and Scharf, 1965). Within these models, the amount of loudness summation is determined by the bandwidth of the critical-band filters and the amount of compression: A higher compression and a smaller filter bandwidth result in a larger loudness summation effect.

In most studies, tone complexes (e.g. Fletcher and Munson, 1933, Zwicker et al., 1957, Scharf, 1959, Scharf, 1961, Cacace and Margolis, 1985, Schneider, 1988, Hübner and Ellermeier, 1993) or bandpass-noise signals (e.g. Zwicker et al., 1957, Port, 1963a) with *long durations* (about 1 s) were used to study this effect. There are only few publications, where temporal aspects of the loudness-summation effect were investigated. Port (1963a) found that the loudness of a 5-msec noise band spectrally centered at 2 kHz increases with increasing bandwidth in the same way as the loudness of a 1200-msec noise band does, given that the absolute loudness of the reference is the same for the two durations. Zwicker (1965) compared the loudness of a 5 kHz tone with that of a white noise for durations of 2 and 500 msec, with the level as the parameter. He found that the level difference necessary to produce equal loudness for the short signals is the same as for the long signals at the same absolute loudness of the references. Both results indicate, that the loudness-summation effect is independent of duration. Scharf (1970) concluded from these results that the critical bandwidth does not change with signal duration. Within loudness models, this conclusion would further indicate that the compression is also independent of duration since both stages determine the loudness summation effect (see above).

However, in two recent studies on temporal summation of loudness it was claimed, that the loudness function must have a different form for short (5-msec) and long (200-msec) tones (Florentine et al., 1996, Buus et al., 1997). The authors based their argument on the finding that the temporal summation of loudness for sinusoidal signals is level-dependent. To obtain equal loudness, the level difference between the short and the long tone increases with increasing level as the level of the short tone varied from 10 to about 70 dB SPL. An increase in level difference implies that the loudness function of the short tone must have a shallower slope than the loudness function for the long tone at the same absolute loudness. A difference in slope results in a different compression and within loudness models this would imply a different loudness summation effect. Thus, loudness summation should be larger for short signals than for the long signals.

On the other hand, there are models of "effective" signal processing in the auditory system (such as, e.g., Dau et al., 1996a+b, 1997a+b), which assume a nonlinear compression behaviour of the auditory system, that incorporates certain time constants ("adaptive compression"). Hence, the transient parts of a signal envelope (i.e., onset and offset of a signal) are less compressed than the stationary parts. In the case of loudness summation, one would therefore expect a smaller effect for short signals than for longer signals. This contrasts with the expectations from the literature described above.

The present study investigates the loudness-summation effect for different signal durations at moderate levels. A constant power level was used for the different durations. This is in contrast to other studies about temporal aspects in loudness summation, where the loudness was held constant and both time *and* amplitude were changed.

3.2 EXPERIMENT 1: LOUDNESS SUMMATION FOR BANDPASS-NOISE SIGNALS AS A FUNCTION OF DURATION

3.2.1 Methods

3.2.1.1 Procedure

Stimuli with different bandwidths were matched in loudness to a reference signal with fixed bandwidth and level using an adaptive 2–interval, 2–alternative forced choice procedure. In each trial the listeners heard two sounds, the reference and the test signal, which were separated by an 500 ms silence interval. Test and reference signal were presented in random order and with equal a priori probability. The listeners indicated which signal was louder by pressing the corresponding key on a keyboard. A simple 1–up 1–down procedure was used, which converges at the 50 % point of the psychometric function (Levitt, 1971). If the listener indicated that the test signal was the louder one, its level was reduced in the next trial, otherwise it was increased. At the beginning, the step size was 8 dB. It was divided by two after each reversal in the adaptive level tracking procedure. At a step size of 2 dB it was held constant for the next 4 reversal. The equal-loudness level for one track was determined by calculating the median of the levels during the last 4 reversals. Three matches were obtained for each listener and pair of stimuli.

To reduce biases that occur when only one stimulus pair was matched in loudness in a series of trials, several interleaved adaptive tracks were used. Hence, concurrent loudness matches were obtained for all stimulus pairs tested (Florentine et al., 1996). On each trial, the track was chosen randomly from all possible tracks, i.e., from all tracks that had not yet been terminated. To ensure that the interleaved tracks converge at roughly the same time the random choice of tracks is further restricted by the following rule: If none of the tracks is terminated, the number of trials for all different tracks has to be the same before the next trial for all tracks is presented in random order to the listener. If one track is terminated, the rule is applied to the choice of trials from the remaining unterminated tracks.

Three different starting levels of the test signal at the beginning of each track were tested for each stimulus pair: 20 dB above, 20 dB below, and at the reference level (0 dB difference). In a series of trials, the tracks very uniformally distributed across the three starting levels. All starting levels had the same frequency of occurrence in a series of trials.

Although the procedure is comparable to the procedure proposed by Florentine et al. (1996) and Buus et al. (1997, 1998), there are some details which differ considerably. First, the additional rule for choosing the next trial was not incorporated in the procedure proposed by Florentine et al. Second, in the present experiment, the reference signal was the same within each series of trials and its level was fixed. It was argued in Florentine et al. (1996) that a fixed reference could bias the measurement. However in the Chapter 2 the fixed reference was shown not to influence the results for the procedure used in this Chapter. Another difference is the choice of the starting level. Whereas in the present paper the tracks started at 20, 0, or -20 dB relative to the reference level, Florentine et al. used a starting level of 10 dB above the expected value. However, at least for the procedure used in the present study the starting level has no significant effect on the results (see Chapter 2).

3.2.1.2 Stimuli and apparatus

The stimuli were generated digitally with a sampling rate of 32 kHz. Stimulus generation and presentation were controlled by a silicon graphics workstation (INDY), which also sampled the listener's responses and controlled the procedure. The software package SI was used for signal generation and control of the experiments, which was developed at the university of Göttingen. The stimuli were D/A converted (16 bits), and then preamplified and lowpass-filtered at 16000 Hz with a computer controlled audiometric amplifier developed in a joint research project on speech audiometry (Kollmeier, 1996). The subjects sat in a double-walled, sound-attenuating booth. The stimuli were presented diotically via a Sennheizer HD 25 headphone without free-field equalization.

Bandpass noise signals with a flat spectrum geometrically centered at 2 kHz were used. To generate the signals, a broadband noise was digitally filtered by setting the magnitude of the Fourier coefficients to zero outside the desired passband. The bandwidths of the test signals were 200, 400, 800, 1600, 3200, and 6400 Hz. The reference bandwidth was 3200 Hz. The level of the reference signal was fixed at 45 dB SPL. Signals were gated with 2.5 ms raised-cosine ramps. Test signals and reference signal had the same duration. Three durations were used: 10, 100, and 1000 ms.

3.2.1.3 Subjects

Ten subjects (6 male, 4 female) participated in the experiment. They ranged in age from 24 to 29. Eight subjects were members of the "Graduiertenkolleg Psychoakustik" or the "Arbeitsgruppe Medizinische Physik" at the university of Oldenburg, one of them was the author JV. The other two subjects (AR, MT) were paid volunteers. Six subjects (CR, JA, JV, PD, KT, OW) had previous experience with equal-loudness judgements. All subjects had normal audiograms (i.e., absolute threshold in quiet $\leq$ 15 dB HL) and no previous history of any hearing problems.

3.2.2 Results

Figure 3.1 shows individual data and standard deviations for 10 subjects. In addition, Fig. 3.2 shows mean data across subjects and interindividual standard deviations. As a function of the bandwidth, the figures show the level difference between the test signal and the reference signal necessary to produce the sensation of equal loudness (equal-loudness level difference). A positive level difference means that, to produce the sensation of equal loudness, the test signal must have a higher level than the reference. The figure shows equal-loudness level differences for signal durations of 1000 ms (squares), 100 ms (circles), and 10 ms (triangles).

In general, all subjects show positive level differences for test-signal bandwidths $<$ 3200 Hz and no level difference when the test signal and the reference signal have the same bandwidth (3200 Hz). For a bandwidth of 6400 Hz, negative level differences are observed. However, there are large individual differences in the shape of the function. Whereas for the subjects KT, OW, and AR equal-loudness level differences for 1000-msec signals decrease continuously with increasing bandwidth, other subjects show an increasing level difference (up to 3 dB) in the range from 200 to 400 Hz and decreasing level differences for bandwidths of the test signal $\geq$ 800 Hz (e.g. subject JV, PD, and SE). Moreover, there are large individual differences among subjects in the amount of loudness summation effect. The level difference between equally loud signals of the same duration with the smallest (200 Hz) and largest (6400 Hz) bandwidth ranges from 10 (subject AR, PD) to 23 dB (subject JV). The range for the maximal difference in level between equally loud signals is even larger (from 10 to 25 dB).

Although large individual differences are apparent, the average data for signals of 1000-msec duration (Fig. 3.2, squared symbols) is consistent with the results presented in the literature for long signals (Zwicker et al., 1957, Port, 1963a, Verhey et al., 1998). The level difference between equally loud signals with the smallest bandwidth and the largest bandwidth is 15.5 dB.

Zwicker et al. (1957) measured loudness summation for 1000-msec bandpass-noise signals at a center frequency of 1420 Hz. In their experiment the level of the test signal was held constant and the level of the reference signal was varied (in contrast to the experiment performed in the present study, where the test signal level was varied). For a test-signal level of 50 dB and a reference bandwidth of 210 Hz, they found a level difference of about 15 dB between equally loud signals with a bandwidth of 210 and 8000 Hz. Port (1963a) compared the loudness of 1200-msec bandpass-noise signals spectrally centered at 2 kHz and presented at 40 dB with a 2 kHz reference tone. He found a level difference of 18 dB between equally loud 200 and 12000 Hz broad bandpass-noise signals.

The individual data for the different durations can be divided into two different groups. For some subjects (CR, JV, and OW) the loudness summation effect is independent of duration. However, the majority of the subjects (70 %) show an increasing loudness summation effect with decreasing duration, i.e., the slope of the equal-loudness level functions increases with decreasing duration. A particularly large temporal effect on loudness summation was found for subject KT. For a signal duration of 1000 ms, this subject showed a level difference of 16 dB between equally loud signals with the smallest and largest bandwidth. The effect amounts 24 dB for 100 ms and 32 dB for 10 ms. Averaged across subjects, the level difference is 15.5 dB for 1000 ms, 18.5 dB for 100 ms, and 23.5 dB for 10 ms.

The increasing interindividual standard deviation with increasing difference between reference bandwidth and the test-signal bandwidth (up to 4.5 dB) reflects the large individual difference in the amount of the loudness-summation effect. Individual listeners are quite consistent in their loudness judgements as indicated by the small standard errors of about 2 dB for most data points in Fig. 3.1. All standard deviations are within 5.5 dB. and their respective amount is independent of bandwidth and duration. This is interesting because most of the subjects reported difficulties to judge the loudness of the short signals and to compare the loudness of long signals when their bandwidths differ considerably. For subjects which showed a duration-dependent loudness-summation effect, the intraindividual standard deviation at the smallest bandwidths (200 Hz and 400 Hz) for durations of 10 and 1000 ms do not overlap. This indicates that at least for this data points, the effect is highly significant.

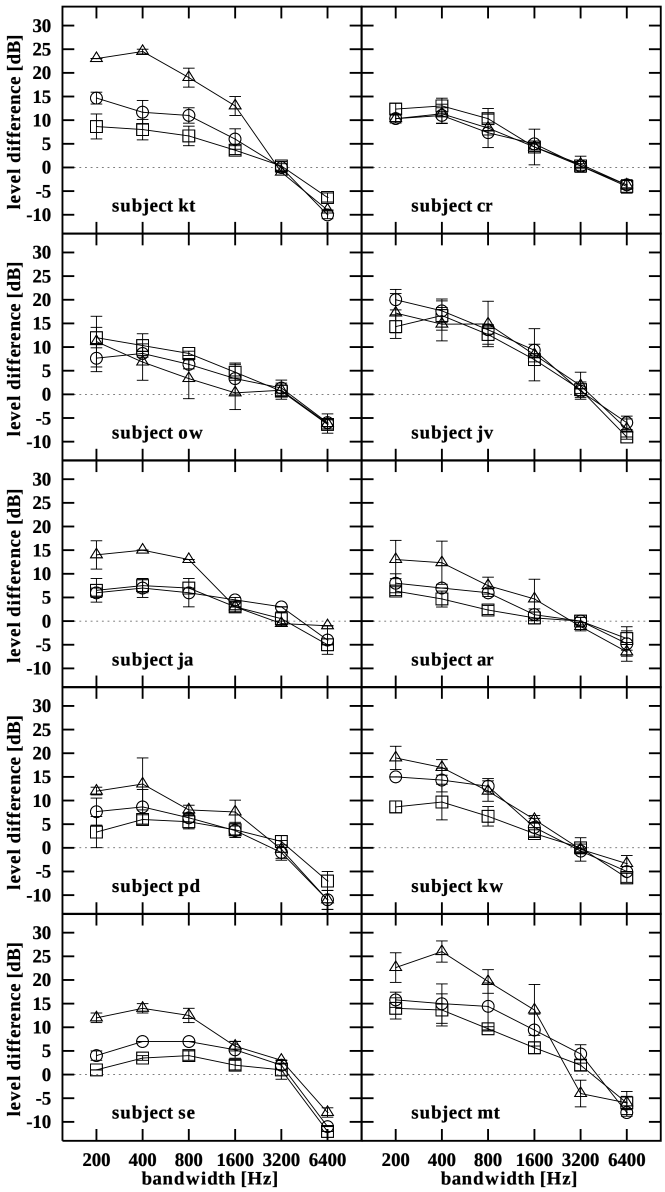

Figure 3.1: *Loudness summation for bandpass-noise signals from ten listeners. Signal duration was either 1000 msec (squares), 100 msec (circles), or 10 msec (triangles). The level differences between the level of test and reference signal needed to obtain equal loudness are plotted as a function of the bandwidth of the test signal. The vertical bars show plus and minus one standard deviation of the mean.*

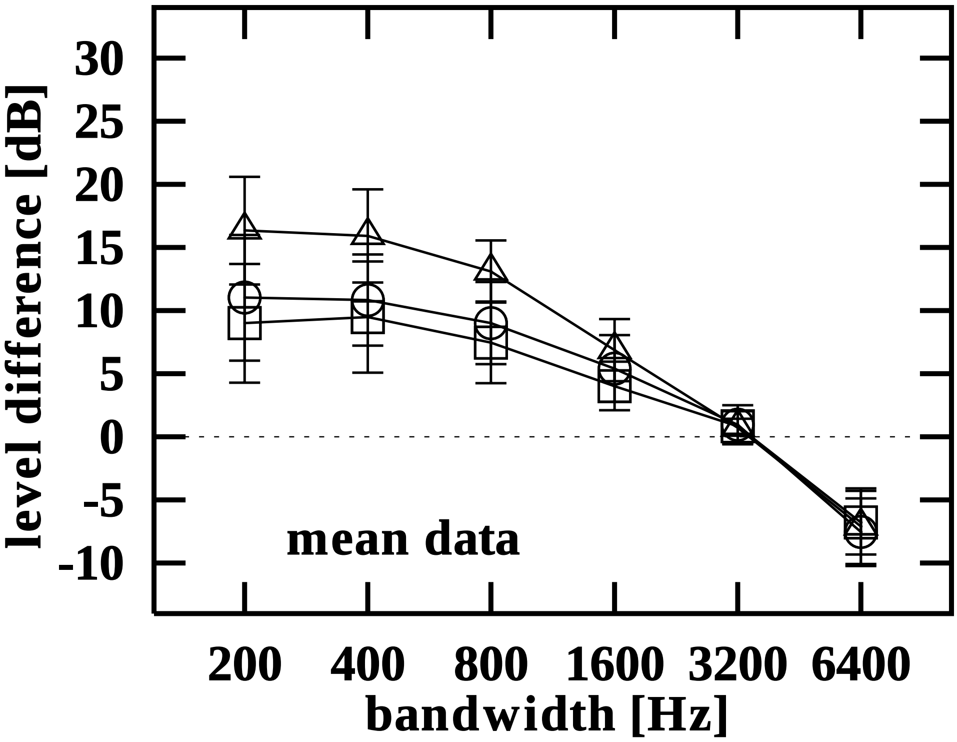

Figure 3.2: *Averaged loudness summation for bandpass-noise signals from ten listeners. Signal duration was either 1000 msec (squares), 100 msec (circles), or 10 msec (triangles). The level differences between the level of test and reference signal needed to obtain equal loudness are plotted in the same manner as Figure 3.1. The vertical bars show plus and minus one interindividual standard deviation from the mean.*

3.3 EXPERIMENT 2: LOUDNESS SUMMATION FOR CHIRP TRAINS AS A FUNCTION OF DURATION

3.3.1 Rationale

Reichardt and Niese (1970) pointed out that for most investigations concerning the loudness of impulsive sounds, tonal signals are preferable as opposed to (narrow) bands of noise, since the necessary filtering of the (broadband) noise source introduces large variations in amplitude. This causes problems if segments of very brief durations are cutted out. The advantage of tonal signals, on the other hand, is a flat envelope and a well-defined temporal structure. It is simple to generate longer signals, where the initial part of a longer signal contains the signal with shorter duration. However, in the present study, tonal signals are not adequate due to their fixed spectral shape. Signals which incorporate both the advantages of tonal signals as well as the possibility to vary the spectrum are frequency modulated signals like chirp trains. They were used in the following experiment to investigate loudness summation for signal durations smaller than 10 ms.

3.3.2 Methods

3.3.2.1 Apparatus and stimuli

The apparatus was the same as in experiment 1. The stimuli were chirp trains with a duration of 2.5, 5, 10, and 40 ms. Test and reference signal had the same duration. They were gated with 0.5 ms raised cosine ramps. For the shortest signal duration (2.5 ms) a single chirp with linearly rising instantaneous frequency was used. To generate longer signals, 2.5 ms long single chirps were concatenated together. To ensure that the time function of the instantaneous frequency changes continuously, each chirp with linearly rising instantaneous frequency was followed by a frequency falling chirp and vice versa. Figure 3.3 shows, as an example, a signal with a bandwidth of 6400 Hz and a duration of 2.5 msec (upper panel) and 10 msec (lower panel), respectively.

The bandwidth of the test signal was 1600, 2263, 3200, 4525, and 6400 Hz[2]. As in experiment 1, the reference signal had a bandwidth of 3200 Hz and a level of 45 dB SPL. Test and reference signal were geometrically centered at 2 kHz.

3.3.2.2 Procedure

As in experiment 1, an interleaved 2–AFC 1-up 1-down procedure was used. In one series of trials, several interleaved adaptive tracks were used to obtain loudness matches at five bandwidths. In contrast to experiment 1, the stimulus pairs in one series of trials had different durations. To keep the number of trials within a block of trials

[2]Since the shortest signals was only 2.5 ms long, bandwidths $<$ 1600 Hz have not been tested.

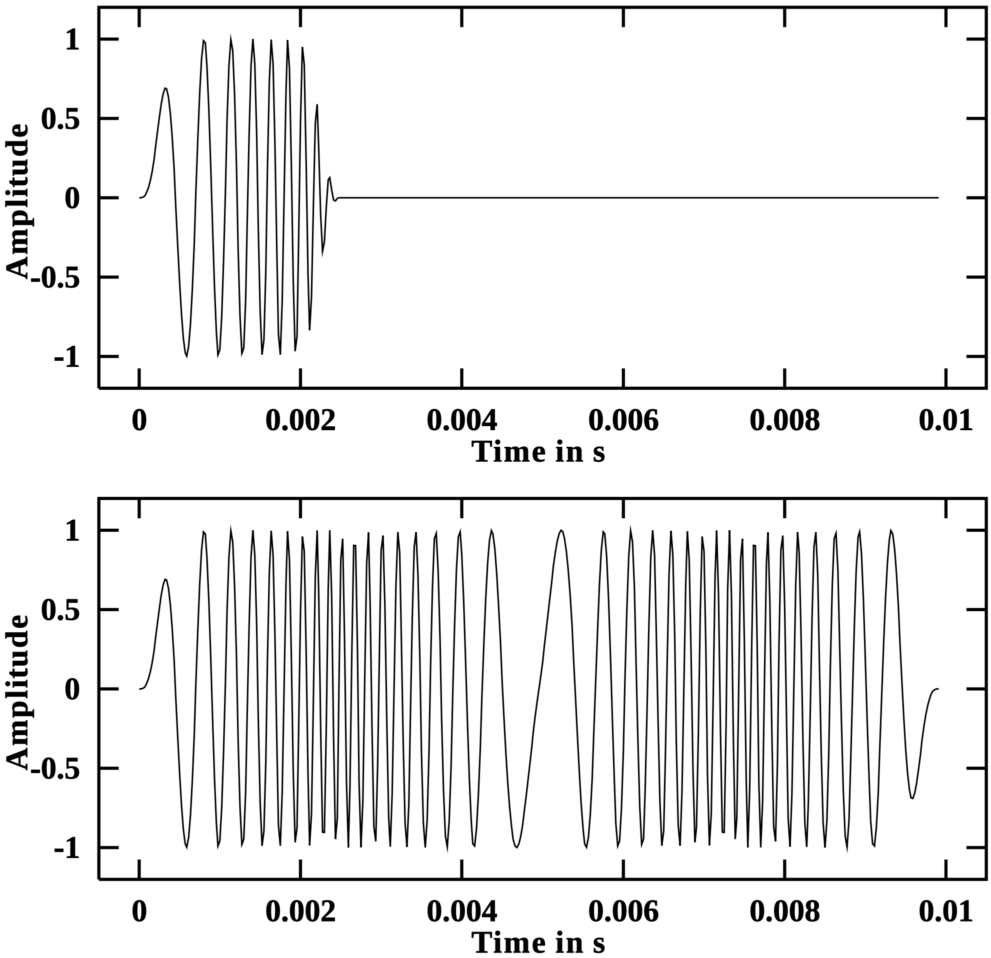

Figure 3.3: *Time signals for chirp stimuli with a bandwidth of 6400 Hz and a duration of 2.5 msec (upper panel) and 10 msec (lower panel), respectively.*

reasonable, the 20 stimulus pairs (five bandwidths and four durations) were divided into two sets. In one set, the test stimuli had the following bandwidth–duration combinations: 1600 Hz–2.5 ms; 3200 Hz–2.5 ms; 6400 Hz–2.5 ms; 2263 Hz–5 ms; 4525 Hz–5 ms; 1600 Hz–10 ms; 3200 Hz–10 ms; 6400 Hz–10 ms; 2263 Hz–40 ms; 4525 Hz–40 ms. The other set included the following bandwidth–duration combination for the test stimuli: 2263 Hz–2.5 ms; 4525 Hz–2.5 ms; 1600 Hz–5 ms; 3200 Hz–5 ms; 6400 Hz–5 ms; 2263 Hz–10 ms; 4525 Hz–10 ms; 1600 Hz–40 ms; 3200 Hz–40 ms; 6400 Hz–40 ms. As in experiment 1, three tracks with different starting levels of the test signal were performed for each stimulus pair.

Subjects

Six normal hearing subjects (4 male, 2 female) ranging in age from 24 to 28 were used. All of them participated in experiment 1.

3.3.3 Results

Figures 3.4 and 3.5 show levels differences to obtain equal loudness as a function of signal bandwidth. Parameter was the signal duration: 2.5 ms (○), 5 ms (□), 10 ms (△), and 40 ms (◇), respectively. Figure 3.4 shows individual data of six subjects and Fig. 3.5 shows the averaged data across these subjects. Because of the extremely short durations of the signals, most subjects reported difficulties to perform the experiment. However, the standard errors of the individual data are as small as those of experiment 1. A difference between experiment 1 and experiment 2 was the subjective quality of the sounds. Whereas the signals in experiment 1 sound noisy, the chirp trains were perceived as buzzing signals with different pitches at different bandwidths.

In agreement with the previous data for bandpass-noise signals, mean equal-loudness level differences (Fig. 3.5) decrease continuously with increasing bandwidth, and assume positive values for test-signal bandwidths smaller than the reference bandwidth and negative values for test-signal bandwidths larger than the reference bandwidth (3200 Hz). On average, the level difference between equally loud 10-msec signals with bandwidths of 1600 Hz and 6400 Hz amounts to only 10 dB compared to 14 dB for bandpass-noise signals (△ in Fig. 3.2). Loudness summation does not depend on signal duration, at least for bandwidths in the range from 2263 to 6400 Hz. For a bandwidth of 1600 Hz, equal-loudness level differences are the same for 2.5, 5, and 10 msec, whereas it is 4 dB higher for 40-msec signals.

Interestingly, the interindividual differences in the amount of loudness summation are small (standard deviations less than 2.5 dB for all bandwidths and durations). For some subjects, the data differ considerably between experiment 1 (where bandpass-noise signals were used) and experiment 2 (where chirp trains were used). In case of 10-msec bandpass noise as the signal, subject KT showed a level difference of 23 dB between equally loud signals with Δf = 1600 Hz and 6400 Hz, in contrast to only 14 dB in case of the chirp trains of the same duration. Similarly, subject JV showed for all durations an effect of 20 dB for bandpass noise and only 7–14 dB for the chirp train.

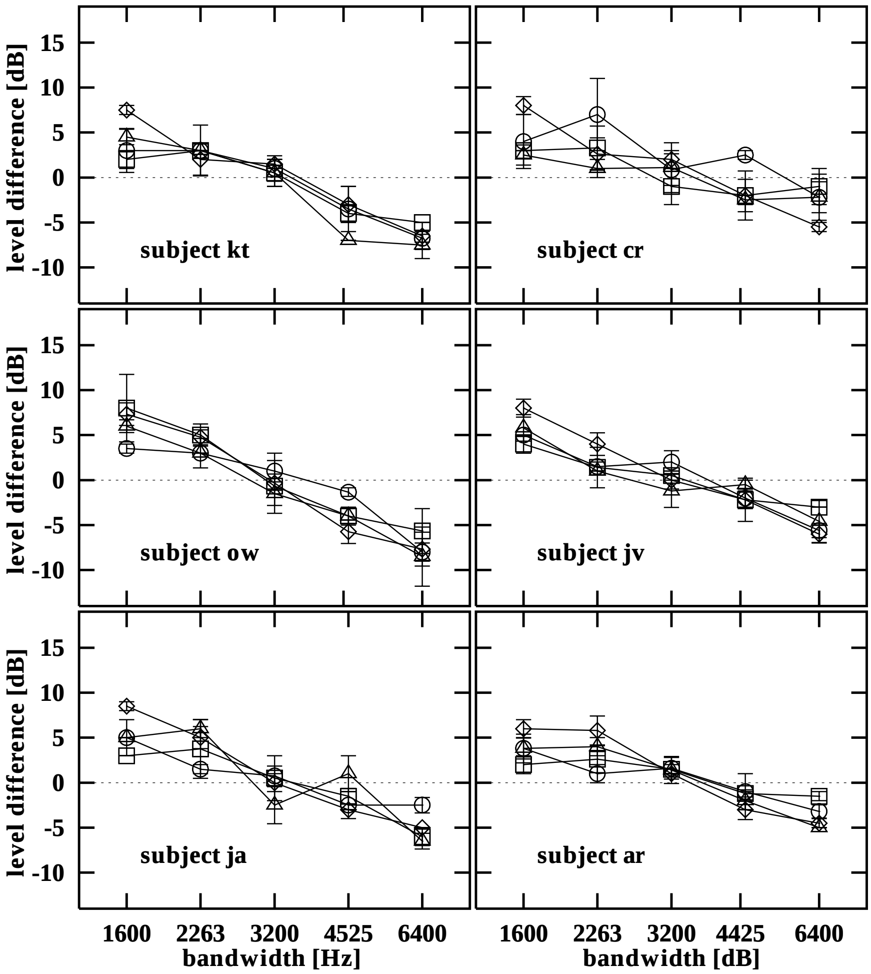

Figure 3.4: *Loudness summation for chirp trains from six listeners. Signal duration was either 2.5 msec (circles), 5 msec (squares), 10 msec (triangles), or 400 ms (diamonds). The level differences between test and reference signal needed to obtain equal loudness are plotted as a function of the bandwidth of the test signal. The vertical bars show plus and minus one standard deviation of the mean.*

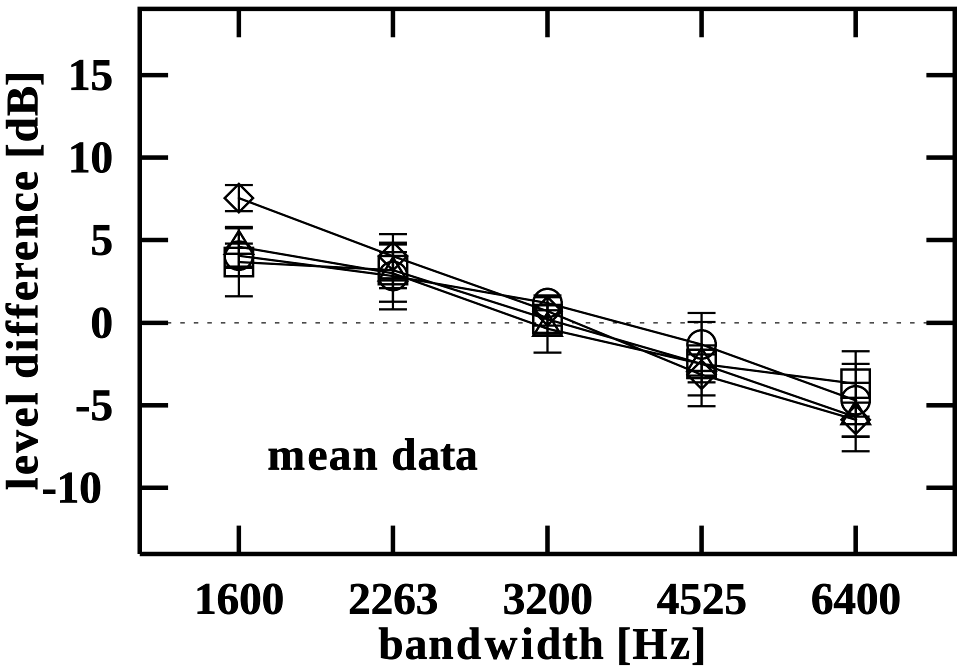

Figure 3.5: *Averaged loudness summation for chirp trains from six listeners. Signal duration was either 2.5 msec (circles), 5 msec (squares), 10 msec (triangles), or 400 ms (diamonds). The level difference between the level of test and reference signal needed to obtain equal loudness are plotted as Figure 3.4. The vertical bars show plus and minus one interindividual standard deviation of the mean.*

3.4 DISCUSSION

3.4.1 Temporal aspects of loudness summation

In contrast to the present study, most of the data described in the literature (Port 1963a, Zwicker, 1965+) did not show an effect of duration on loudness summation (see introduction). There is only one study in which a duration-dependent loudness-summation effect was found (Boone, 1973). As the reference signal, Boone (1973) used a 1 kHz tone at a level of 72 phon. The test signals was a bandpass noises at various bandwidths with a flat spectrum geometrically centered at 1 kHz. Boone found an increasing loudness summation effect with decreasing signal duration. To produce equal loudness for the reference tone and a test signal with a bandwidth of 1800 Hz, he measured a level difference of 16 dB for a duration T of 18 ms whereas he obtained only 9 dB for a duration of 1000 ms. This finding that the level difference for short-duration signals is twice the level difference for long-duration signals is in line with the experimental results of the first experiment in the present study[3]. Interestingly, Boone discussed only the relationship between bandwidth and duration. He discussed critically the critical bandwidth of 500 Hz derived from his data at a center frequency of 1 kHz. In his opinion, an unacceptable spread between the sixteen test subjects was the reason for this unexpected large critical bandwidth. This might explain why he did not discuss the large discrepancy in the amount of loudness summation between his results and those of other authors.

Boone (1973) and Port (1963a) measured loudness summation for different durations at nearly the same loudness, whereas in the present study the reference level was kept constant for the different durations. Therefore, a direct comparison might be problematic. However, the difference in loudness level of a 10-msec signal and a 1000-msec signal with the same level (in dB SPL) amounts to only 10 phon (e.g. Zwicker and Fastl, 1990). For stationary signals, Zwicker et al. (1957) showed that loudness summation decreases for loudness levels smaller than about 50 phon. The present data were obtained at about 55 phon for the long-duration signal. Assuming a duration-independent loudness summation effect at the same absolute loudness (in phon) , loudness summation should be the same or slightly smaller for the 10-msec signal compared to 1000 msec signals, which contrasts with the findings in experiment 1. Hence, the duration dependence of loudness summation found here can not be explained on the basis of level-dependent loudness summation.

Since Zwicker (1965) measured loudness summation for a wide range of levels, loudness summation can also be compared in his data at the same physical level. To obtain equal loudness for a 5 kHz tone-pulse at 55 dB SPL and a white-noise pulse at various durations of both signals, he obtained a 5 dB smaller level difference for 2-msec long signals compared to 500 msec long signals. Our own results show the opposite effect. It

[3]To compare the present data with Boone's data, we have to transform the absolute bandwidths in Fig. 3.2 to bandwidths relative to the center frequency CF. On this scale we obtain a level difference of about 17 dB (T = 10 ms) and 9.5 dB (T = 1000 ms) between a signal with a subcritical bandwidth ($\Delta f/CF < 0.2$) and a signal with a relative bandwidth $\Delta f/CF = 1.8$ (i.e., an absolute bandwidth of 3600 Hz at a center frequency of 2 kHz)

is unclear why such a large difference between our results and the results from Zwicker occurs. The reduced loudness summation in Zwicker's experiment could be explained by a spread of excitation for the extremely short duration (2 ms), which reduces the spectral difference between the tone pulse and the white-noise pulse. In addition, the differences in the experimental procedures might contribute to the different results. It was shown in the Chapter 2 that the use of a simple adaptive procedure might produce a difference in loudness summation for short and long signals of 1 to 4 dB, depending on the starting level of the test signal.

3.4.2 Temporal integration of loudness and loudness models for impulsive sounds

Most studies investigated temporal integration of loudness without changing the signal spectrum explicitly. Usually tone bursts were used as the signal (Munson, 1947, Poulsen, 1981, Namba et al., 1987, Tachibana et al., 1987, Kumagai et al., 1982b), but there are also studies where temporal integration of loudness for burst of more complex sounds as e.g. multitonal complexes (Zwicker, 1966, Scharf, 1959, 1961), narrowband noise (Port, 1963a+b) and broadband noise (Port, 1963a+b, Kumagai et al., 1982b, Florentine et al., 1996, Buus et al., 1997) were investigated.

In experiments on temporal integration of loudness it was found that loudness increases linearly with time for durations up to a "critical duration" of about 100 ms. For longer durations loudness is independent of duration. According to Port (1963a), the loudness level for signals of less than the critical duration is primely determined by the energy (integral of sound intensity over the duration of signal) of the signal, whereas it is determined by stimulus intensity itself for longer signals. Within simple models, the build-up of loudness is described by two stages: A square law rectifier and a single RC-network, i.e., the build-up of loudness is assumed to be governed by a single time constant τ_i (e.g. Niese, 1959, Reichardt, 1965, 1970). There is considerable confusion in the literature as to the exact numerical value of the time constant. In Niese's model (1959) the time constant was 25 ms, Port (1963a) found a time constant τ_i = 70 ms, Zwicker (1966a) proposed a time constant of 150 ms. Sone et al. (1986) concluded, that a time constant of 125 ms to 1 s is the most suitable for the evaluation of stimuli used in their study. Takeshima et al. (1988) reported about a growth of loudness even for durations beyond 500 ms, indicating that the time constant might be even larger than that proposed by Zwicker (1966a). Reichardt (1965) pointed out, that there are large interindividual differences with respect to the temporal integration of loudness. He concluded that the use of only few subjects in most studies might be the reason for the different time constants. Measuring loudness as a function of tone-impulse duration, Reichardt and Niese (1970) showed that subdividing their group of 50 subjects in subgroups of 6, the time constant from the different subgroups can vary between 30 ms and 100 ms, whereas mean data over all (50) subjects yields a time constant of 50 ms.

Poulsen (1981) reported in his investigations of temporal integration of loudness of tone pulses about discrepancies between predictions of a single-time-constant model and his experimental results obtained with 25 subjects using pulses of durations less than

20 ms. To account for the data, he proposed a combination of two time constants, one about 5 to 10 ms and the other one about 100 ms. Poulsen also pointed out that the data from 300 subjects of a Round Robin Test on impulsive noise (Pedersen et al., 1977), can be predicted by a two-time-constant model, whereas a single-time-constant model only accounts for data at very high signal levels of 95 dB. A loudness model with two exponentially averaging circuits, one with a rise time of 5 ms, and the other with a rise time of 125 ms, was also proposed by Kumagai et al. (1982a+b, 1984).

Based on measurements of loudness of repeated burst, some authors demanded a separate decay-time constant τ_d, which should be larger than the rise-time constant (Port, 1963a, Kumagai et al., 1984, Ogura et al., 1991, 1993). As in the case of τ_i, many different values were proposed for τ_d in the range from 250 ms (Kumagai et al., 1984) to 5000 ms (Ogura et al., 1991, 1993). Reichardt and Niese argued that the increased decay-time constant found in loudness measurements with pulse trains might be a consequence of the additional roughness component yielding an increase in the annoyance value and is not a true loudness effect (Niese, 1965, Reichardt and Niese, 1965, Reichardt, 1970). However, because a decay-time constants τ_d only plays a role for loudness measurements with repeated bursts, it is not relevant for the results of the present study, where the loudness of single bursts was obtained.

Although some results indicated that the temporal integration of loudness depends on physical parameters of the signal such as center frequency (Boone, 1973), bandwidth (Port, 1963a+b) and level (Poulsen, 1981), most loudness models assume a single overall temporal weighting (e.g. Port 1963b, Ogura et al., 1993, Stone et al., 1997) probably for reasons of simplicity. Within these models, spectral effects are described by spectral analysis (critical-band filtering, compression, summation across critical bands) before temporal integration. It is assumed that the spectral analysis, which is responsible for loudness summation, works nearly instantaneously (e.g. Ogura et al., 1993), whereas the overall temporal weighting has longer time constants of about 100 ms as discussed above. Such a model structure would predict a loudness-summation effect, which is independent of signal duration, in contrast to the present results (experiment 1)[4].

A more complex model for calculating loudness of temporally variable sounds was proposed by Zwicker and Fastl (1990). In contrast to the model described above, they assumed a duration-dependent lowpass filtering within each critical band before summation across critical bands followed by an overall time weighting with a lowpass filter (τ about 100 ms). They based their assumption of a non-instantaneous spectral-analysis stage on results of experiments with strongly frequency modulated sounds (Zwicker, 1974, 1977) and on experiments where the influence of temporal and spectral structures of sequences of tone bursts on loudness was investigated (Zwicker, 1969). The lowpass filter within each critical band had a short rise time (few milliseconds) and a long decay

[4]Interestingly, loudness summation for different durations was usually measured at the same absolute loudness level, i.e. the level of the short duration reference was higher than the level of the long-duration reference. However, because loudness summation is determined by the spectral analysis within a model, models with a nearly instantaneously acting spectral-analysis stage would predict the same loudness summation for different durations, when the *level* of the references is the same for the different durations. The results of experiment 1 (measured at the same reference level!) contradict this prediction.

time. The decay is determined by experiments on forward masking. However, this model cannot account for the the data in experiment 1, since only the decay time varies with duration, which is not important for the experiments in the present study (see above).

3.4.3 Further evidence for a duration-dependent loudness summation

The combination of experimental results on temporal integration of loudness for different bandwidths with results on loudness summation for stationary signals indirectly gives information about temporal aspects of loudness summation. One can construct equal-loudness level surfaces where one axis represents duration, one represents bandwidth, and the other axis represents the level. On such an equal-loudness level surface, level differences for equally loud signals can be estimated that have not directly been measured (e.g., short-duration signals with different bandwidths).

In this way, it is possible to extract information about temporal aspects of loudness summation from the data presented by Florentine et al. (1996). They measured the level difference between equally loud 5- and 200-msec stimuli (1 kHz-tone or white noise) for a wide range of levels. Combining their results with results on loudness summation for stationary signals presented by Zwicker et al. (1957), we can derive loudness summation for short signals at different levels. For a 5-msec white-noise signal with a level of 50 dB, the level of an equally loud 5-msec sinusoidal signal should be 18 dB higher (see Appendix A for a computation of the level difference). The loudness-summation effect is by 6 dB larger than for long signals (12 dB). The same loudness summation effect is obtained for a level of the short noise signal of 40 dB. The increase in loudness summation effect for short signals derived from Florentine et al.'s data at moderate levels agrees with the results of the present study.

Interestingly, the difference in loudness summation for short and long signals derived from Florentine's data decreases with increasing level. For a level of 70 dB for the short noise signal, loudness summation for short signals is only 2 dB larger than loudness summation for long signals. Finally, for $L_{short}(noise) = 90$ dB, the loudness summation for short signals is even 3 dB *smaller* than the loudness summation for long signals. However, experiments at such high levels are problematic. To decide how the difference of loudness summation for short and long signals changes with level, it is desirable to perform experiments where the level dependence of spectral integration of loudness for different durations is tested explicitly.

3.4.4 Relation to physiological data

The average firing rate of an auditory nerve fiber has a large initial peak at stimulus onset (onset response) followed by a slow decay to a lower sustained firing rate (steady state response). The decay can be described by two exponential components, rapid adaptation with a time constant of a few milliseconds and short-term adaptation with a time constant of about 50 ms (e.g. Smith, 1988, Yates et al., 1985, Smith and Zwislocki, 1975). A different compression for short and long signals (as suggested to account for the findings in experiment 1) would imply that the onset response changes in a different way with level than the steady-state response does . In fact, the ratio between onset response and steady-state response should decrease with level, when the compression is higher for short signals than for long signals. Unfortunately, the literature is ambiguous with respect to this point. Müller and Robertson (1991) reported differing behaviour for different types of auditory nerve fibers. They found an increasing ratio for increasing adapted firing rate (steady state response) for high-spontaneous rate fibers, whereas the ratio decreases for low-spontaneous rate fibers. For medium-spontaneous rate fibers, the ratio is nearly independent of the adapted firing rate. Several authors (e.g. Smith, 1988) reported that the ratio depends on the time interval used to calculate the onset rate. For a time interval of 1 msec, the ratio between onset rate and steady-state rate is independent of the level and increases with the level for high intensities. In contrast, the onset rate calculated over 5 msec always increases in proportion to the steady state response (Smith, 1988). However, Smith and Zwislocki (1975) reported that onset rate measured in a time interval of 10 ms tended to be smaller than required for a constant ratio for high intensity levels (15 dB above threshold of the fiber).

It is generally assumed, that several auditory nerve fibers are combined to encode loudness information. However, the general mechanism is still unclear. Relkin and Doucet (1997) deduced from their recordings of perstimulus compound action potential for tone bursts with different center frequencies that a simple sum of the spike activity in the auditory nerve cannot be the physiological correlate of loudness.

Summarizing, some data for auditory nerve fibers indicate a different compression for short signals as for long signals, others do not. Furthermore, a general problem is that it is still unclear how the information carried by the auditory nerve fibers is combined to create the loudness percept.

3.4.5 Implications for dynamic loudness models

As already stated above, loudness models with an instantaneously acting spectral-analysis stage followed by an overall temporal weighting do not account for duration-dependent effects of loudness summation. One possibility to describe the results of the present study is to include a duration-dependent compression in the spectral analysis stage. As mentioned above, the difference in compression for long and short-duration tones might also vary with level.

To incorporate a duration-dependent compression, an adaptation stage at the output of each critical band might be used. Such a model structure was implemented in the signal processing model by Dau et al. (1996a+b, 1997a+b), which was developed to account for data of detection and masking experiments. So far, the model was not extended for modeling loudness perception. However, the adaptation stage within the model transforms fast input variations linearly whereas stationary signals are transformed almost logarithmically. This implies an increasing onset/steady-state ratio with increasing level. As a consequence, a loudness model incorporating such an adaptation stage would predict a *smaller* loudness-summation effect for short signals than for long signals, in contrast to the data. To account for the present data an adaptation stage is needed where the onset-to-steady-state ratio decreases with increasing level.

An alternative to simulate a duration-dependent loudness summation would be to assume temporal integration within each critical band, where the time constants depend on the center frequency. Assuming that the time constant of the additional excited critical bands decrease with increasing bandwidth, then the contribution of the additional excited critical bands will be higher when short signals are applied than in the case of stationary signals. As a consequence, loudness will increase faster with increasing bandwidth than for stationary signals. However, experimental data for temporal integration of loudness do no directly support such an approach (e.g. Port, 1963a, Poulsen, 1981).

In the introduction, it was noted that – in addition to compression – also frequency selectivity will affect loudness summation. Therefore, a further approach to account for the data in experiment 1 would be the assumption of time-varying peripheral filter shapes. In fact, selectivity should be high at the beginning of a signal and should decrease with increasing delay to the onset of the signal. However, experiments on the temporal effects in spectral masking do not support such a hypothesis. Moore et al. (1987) performed notched-noise experiments with a short tone presented at different delays relative to the onset of the masker. In agreement with the conclusion drawn by Zwicker and Fastl (1972) in a review article about temporal development of frequency selectivity, they stated that frequency selectivity does not develop over time, but is present almost instantaneously. Tone-on-tone masking experiments sometimes even show an *increase* in frequency selectivity over time (e.g. Bacon and Viemeister, 1985, Bacon and Moore, 1986a+b). However, Moore et al. (1987) pointed out that it is still unclear, whether the filtering process itself, or some higher process following the filtering, is time varying. Such a higher process, but operating in the opposite direction, might also yield an increased frequency selectivity for loudness perception at the onset of a signal. However, it is more likely that compression is time-varying.

3.5 SUMMARY

Loudness summation depends on signal duration. It is markedly increased for 10-msec signal, whereas it is almost the same for 100-msec and 1000-msec signals. There seems to be no further increase in loudness summation for signal durations smaller than 10 msec. Present loudness models of time varying sounds cannot account for the experimental data. The reason for this failure is that their spectral-analysis stage acts nearly instantaneously and duration-independent. A model which contains a duration-dependent compression should in principle account for the present data. Further studies should be done to test, how the temporal effects in loudness-summation depend on the level.

Chapter 4

Simulations of spectral masking and the asymmetry of masking

ABSTRACT

This study presents measurements and simulations of spectral masking. Three different experimental conditions were examined: (i) Notched-noise masking with a test signal spectrally centered in the notch, (ii) notched-noise masking with an asymmetric notch around the test-signal frequency, and (iii) masking experiments where the test signal and the masker have the same center frequency but different bandwidths. The simulations were performed with a model originally developed to describe effects of temporal masking and modulation detection [Dau et al., J. Acoust. Soc. Am. 102, 2892–2905 (1997)]. The model accounts well for notched-noise data which mainly is a consequence of the gammatone-filter parameters used in the peripheral filter stage of the model. The model also accounts for effects of spectral masking in off-frequency conditions of the second experiment, which results from the decision stage that integrates signal information across frequency. A simple modulation-lowpass filter (energy integrator) would be sufficient to describe these experimental results. In contrast, the processing by a modulation filterbank is necessary to account for the conditions of "asymmetry of masking" in the third experiment, where it is observed that thresholds for test signals with bandwidths larger than the masker bandwidth are much lower than those for the reversed condition. In conditions, where the test-signal bandwidth is larger than the masker bandwidth, the model is able to use the inherent higher modulation frequencies of the test signal as an additional cue. A model which only processes the energy of the signal fails to predict the results.

4.1 INTRODUCTION

Fletcher (1940) measured the detection threshold of a sinusoidal signals in the presence of a bandpass-noise masker as a function of the bandwidth of the masker. He concluded from his results that (i) the peripheral auditory system consists of a bank of overlapping critical-band filters, and (ii) the thresholds are determined by the overall energy of the masker in the peripheral filter centered at the signal frequency. Based on these assumptions, Fletcher proposed a model known as the power-spectrum model, which works well in many situations. In the original version of this model, rectangular peripheral filter shapes were assumed. However, results from notched-noise masking experiments (e.g., Patterson, 1976, Patterson and Nimmo-Smith, 1980) showed that the critical bands have shallower slopes. Patterson and Nimmo-Smith (1980) described their data with a power-spectrum model using rounded-exponential (roex) filters, where the filter function decays nearly exponentially at each side of the center frequency of the filter. Later on, Patterson et al. (1987) proposed the gammatone filterbank as a powerful computational approximation of the roex filters.

In addition, Patterson and Nimmo-Smith (1980) compared the results of the above experiments, where a sinusoidal signal is masked by a notched noise with a symmetric notch, with masking patterns, where a lowpass noise was used as the masker. The spectral distance between the signal frequency and the cut-off frequency of the lowpass masker was the same as the distance between the signal and the edges of the noise in the notched-noise condition. Compared to the notched-noise condition, in the lowpass-noise condition up to 10 dB lower thresholds are measured. Using the same (symmetric) peripheral filter as in the notched-noise condition, centered at the signal frequency, a power spectrum model would predict only a 3 dB difference between notched-noise and lowpass-noise condition, since only half of the masker energy would fall in the peripheral filter in the lowpass-noise condition. To account for the large threshold difference of 10 dB between the two conditions, Patterson (1976) proposed that the subjects use information from a peripheral filter above the test-signal frequency (off-frequency listening) in the lowpass-masker situation. Patterson and Nimmo-Smith (1980) showed, that a power-spectrum model using peripheral filters at different center frequencies in both conditions indeed can reproduce the large threshold difference between the two conditions.

However, there are a number of experimental results that can not be explained by the power-spectrum model. For example, thresholds can decrease markedly when the masker is modulated , as observed in comodulation-masking-release (CMR) experiments (e.g. Hall et al., 1984). In the literature it is generally assumed that the release is a consequence of an across-channel process, where the subjects make use of the coherent envelope fluctuations of the masker in peripheral filters at different center frequencies. However, it is also argued in the literature (e.g., Schoonefeldt and Moore, 1989), that some of the release might be due to a single-channel process, where the threshold is determined by a change of the amplitude statistics within the critical band centered at the signal frequency. Verhey et al. (1998, see also Chapter 5) showed that the reduced thresholds for a modulated masker in some CMR-condition can be predicted if high modulation rates are processed (by a modulation filterbank) in addition to the signal energy, whereas a simple energy detector would fail to predict CMR.

Another phenomenon , which can not be described by the power-spectrum model is related to the differing masking properties of a bandpass noise and a sinusoid. Hellman (1972) found that a bandpass-noise signal with a subcritical bandwidth is audible at signal-to-masker ratios as low as -20dB to -30 dB, when it is masked by a sinusoidal signal at the center frequency of the noise signal. In the reversed condition, with a sinusoidal test signal and a bandpass-noise masker, the test signal is audible at signal to masker ratios as high as -4 dB. Hall (1997) investigated this asymmetry of masking for a set of test-signal and masker bandwidth combinations. The signals had a center frequency of 1 kHz. The bandwidth ranged from 0 to 256 Hz for the test signal and the masker. Hall (1997) found that for the conditions where the test-signal bandwidths are larger than the masker bandwidth, thresholds decrease markedly with increasing signal bandwidth, whereas for test-signal bandwidths equal or smaller the masker bandwidth, thresholds are essentially constant. Hall (1997) concluded that the detection is based on the long-term average energy as long as the bandwidth is equal or smaller the masker bandwidth, but that other cues are utilized when the test-signal bandwidth exceeds the masker bandwidth. He suggested that a model which operates on the temporal structure of the signals may be able to reproduce his results. However, up to now, no simulations with such a model were performed to test this hypothesis.

In the present study, simulations of spectral masking are compared with corresponding own experimental data. This simulations are based on the modulation-filterbank model proposed by Dau et al. (1997a+b). First, the ability of the model to account for the frequency selectivity of the auditory system in "classical" notched-noise conditions and to account for the masking patterns in typical off-frequency conditions is tested. Then, the experimental conditions described by Hall (1997) are replicated here using running noise for the test signal and the masker as well[1]. Several implementations of the model are tested to clarify the role of off-frequency listening and temporal effects in spectral masking.

[1]Running noise is used to avoid that the subjects may focus on the time structure of a certain realization of the signals.

4.2 MODEL STRUCTURE

Dau et al. (1996a+b, 1997a+b) presented a model of signal processing in the auditory system which was able to describe a large number of detection and masking experiments. The model combines several preprocessing stages with an optimal detector as the decision device (see Fig. 4.1). The first stage of the model is the linear gammatone-filterbank model of Patterson et al. (1987). The signal at the output of the filters is halfwave rectified and lowpass filtered at 1000 Hz. Thus, at high center frequencies only the signal envelope is further processed. A chain of five consecutive nonlinear feedback loops is incorporated (Püschel, 1988, Dau et al., 1996a) to account for adaptation and compression in the auditory system. A linear modulation filterbank further analyzes the amplitude fluctuations of the envelope. To model the limits of resolution, an internal noise with a constant variance was added to the output of each modulation filter output. In the decision process, a stored temporal (internal) representation of the signal to be detected (the template) is compared with the actual activity pattern. This is done by calculating the cross correlation between the two temporal patterns. This procedure is comparable to a "matched filtering" process.

In the present study a slightly modified version of the original model is used which was proposed by Verhey et al. (1998, see also Chapter 5): The center frequency of the highest modulation filter is set to a quarter of the center frequency of the peripheral filter. In the original version it was 1000 Hz. This slight modification was motivated by physiological findings by Langner and Schreiner (1988). Apart from that, all parameters of the model were identical to those described by Dau et al. (1997a+b). In an earlier version of the model (Dau et al., 1996a+b), a single modulation-lowpass filter with a cut-off frequency of 8 Hz was applied instead of a modulation filterbank. It was shown that such a model structure accounts for masking effects such as forward masking and the detection of test tones in noise. However, the complete modulation-filterbank model is required to quantitatively model findings in modulation detection and masking experiments (Dau et al., 1997a+b)

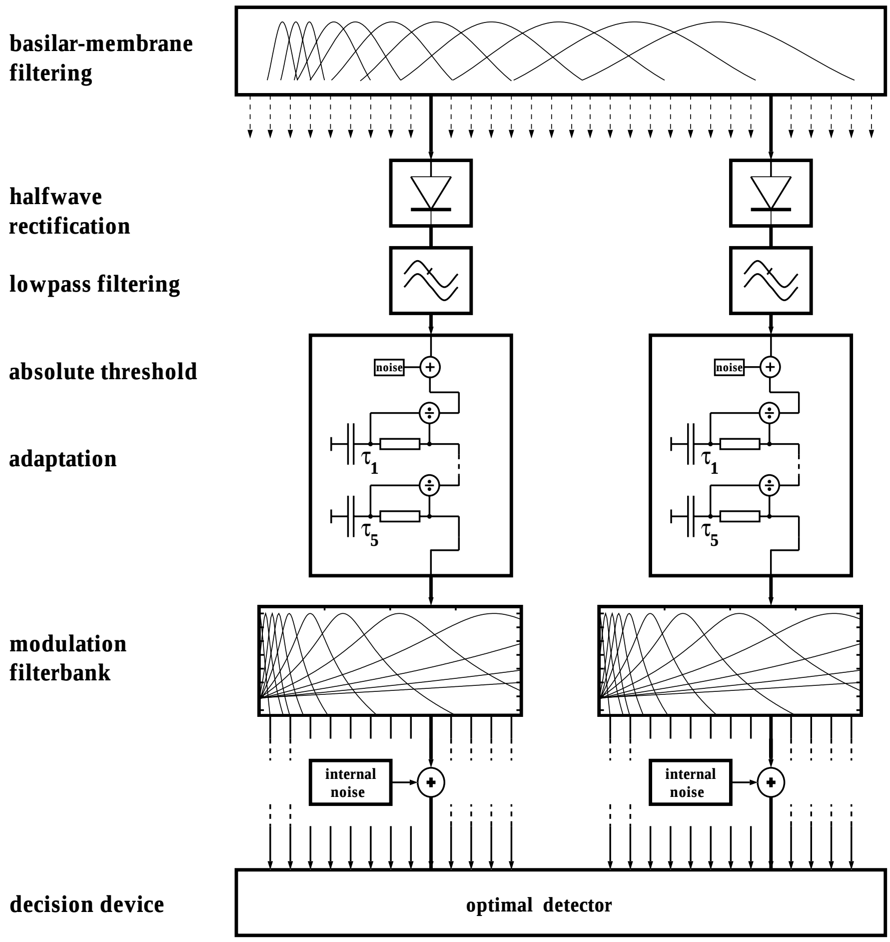

Figure 4.1: *Block diagram of the psychoacoustical model as described in Dau et al. (1997a+b). The signals are preprocessed with a gammatone filterbank, subjected to adaptation and filtered by a modulation filterbank in each peripheral filter. Finally an internal noise is added. The decision device is realized as an optimal detector.*

4.3 EXPERIMENT A: Spectral masking with a symmetric notched-noise masker

4.3.1 Methods

4.3.1.1 Procedure

A 3-interval forced-choice procedure (3 IFC) was used to measure the masked threshold of a noise signal in the presence of a noise masker. A trial contained three intervals separated by 500 ms of silence. Two intervals consisted of the masker alone, and one randomly chosen interval contained the masker plus signal. The subjects task was to indicate the signal interval by depressing the corresponding key on a keyboard. Feedback was provided on a screen in front of the subjects as to whether the response was correct or not. The level of the signal was varied according to a 1-up 2-down procedure. This procedure estimates the 70.7 % point on the psychometric function (Levitt, 1971). The step size was 8 dB at the beginning of the experiment. It was reduced to 4 dB after the first reversal and finally to 2 dB after the second reversal. Then the step size was held constant over the next 10 reversals. The threshold was determined by calculating the median of the levels during the last 10 reversals. The final threshold estimate was taken as the mean over three runs.

4.3.1.2 Stimuli and apparatus

A broadband noise masker with a spectral notch at 1 kHz was used. The masker was generated by setting the fourier components to zero inside the desired notch width (FFT length: 96000 samples, sampling frequency 32 kHz). The notch width Δ f was 0 (i.e., no notch), 100, 300, and 500 Hz, respectively. Masker duration was 300 ms. This period was randomly cut out from a fixed noise waveform which was generated at the beginning of each experiment and had a periodicity of 3 s. 10 ms cosine ramps were applied to the masker. The spectrum level of the noise masker was 40 dB. The test signal was a 100-msec long 1-kHz tone. The test signal was temporally centered in the masker and was gated with 10 ms raised-cosine ramps.

The stimuli were generated digitally with a sampling rate of 32 kHz. Stimulus generation and presentation were controlled by a silicon graphics workstation (INDY), which also sampled the listener's responses and controlled the procedure. The software package SI developed at the university of Göttingen was used for signal generation and controlled the experiments. The stimuli were converted into analog signals with the on-board 16-bit D/A converter, and then preamplified and lowpass-filtered at 16000 Hz with a computer controlled audiometric amplifier developed in a joint research project on speech audiometry (Kollmeier, 1996). The subjects sat in a double-walled, sound-attenuating booth. The stimuli were presented monaurally via a Sennheizer HD 25 headphone.

4.3.1.3 Subjects

Four male normal-hearing subjects (absolute threshold in quiet $\leq$ 15 dB HL) participated in the experiment, two of them being the authors TD and JV. All subjects had normal hearing and no previous history of any hearing problems. The subjects were between 26 and 33 years old. The authors had extensive prior experience in performing spectral masking experiments. The other two subjects (JA, PH) were inexperienced listeners.

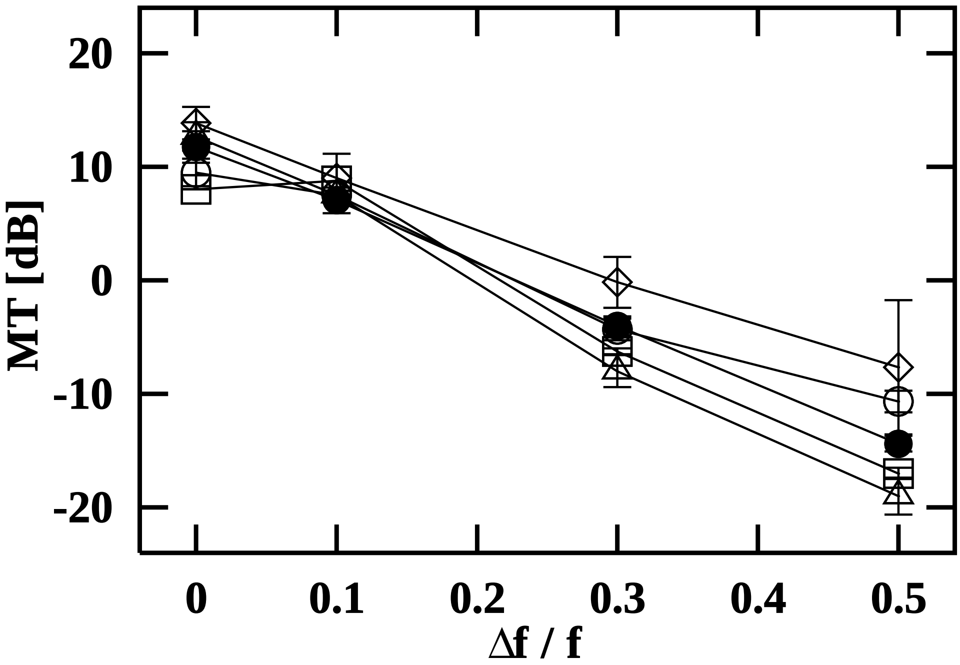

Figure 4.2: *Masked thresholds (MT) for a sinusoidal signal presented in notched noise as a function of masker bandwidth. The signal frequency was 1 kHz. Open symbols indicate measured thresholds (JA (○), PH (◇), TD (□) JV (△)). Closed symbols indicate simulated thresholds.*

4.3.2 Results

Figure 4.2 shows individual data and intraindividual standard deviations for 4 subjects. As a function of the notch width, the figure shows masked thresholds (in dB relative to the spectrum level of the masker) of a sinusoidal signal in the presence of a notched-noise masker. The notch width is given as a fraction of the center frequency of the notch. Open symbols indicate measured data of four subjects. As expected from the data in the literature, thresholds decrease with increasing notch width. Clear individual differences are apparent in the measured data. Differences in threshold of 2 to 12 dB across subjects can be observed. This, however, is consistent with the results from the literature (e.g., Moore et.al., 1995)

Simulated data are indicated by the closed symbols. The simulations were performed with a model, which is identical to the model version used by Verhey et al. (1998) to model within-channel cues in comodulation masking release (cf., Chapter 5). This model processes the output of only one peripheral channel centered at the test-signal frequency and uses a modulation filterbank to analyses the amplitude fluctuations. The general shape of the simulated threshold curves coincides with the measured curves. The difference between simulated and individually measured data maximally amounts to 5 dB. Note, however, that a standard peripheral filter bandwidth is used in the model, i.e., the bandwidth is not adjusted individually. In fact, the difference between the simulated data and the mean measured data (not shown) is less than 2 dB.

To test the contribution of high modulation frequencies to the simulated thresholds, a single modulation-lowpass filter with a cut-off frequency of 8 Hz as proposed by Dau (1996a+b) was used instead of the modulation filterbank. The simulated results of the two model versions do not differ markedly, indicating that a modulation filters at higher frequencies do not contribute important information in this kind of experiments (not shown).

In addition, the possible role of off-frequency information was investigated using the multi-channel version of the model, where the information of several peripheral channels around the test-signal frequency was combined. The envelope of each peripheral filter was processed with a modulation-lowpass filter. The detectability of the signal was not improved within the model when off-frequency channels were considered (not shown). This finding indicates that mainly the energy of the signal in the peripheral filter tuned to the test-signal frequency accounts for the experimental results in notched-noise experiment with a symmetric notch. In addition to the experiment described in this section, it is shown in Appendix B that the model also accounts for effects of spectral masking in different frequency regions.

4.4 EXPERIMENT B: Spectral masking with an asymmetric notched-noise masker

4.4.1 Methods

As in experiment A, a 3-IFC 2-down 1-up procedure was used. The apparatus differed slightly from the apparatus used in the first experiment. Signal generation and presentation were controlled by a SUN-Workstation and the stimuli were transformed to analog signals with the aid of an external 16-bit D/A converter (Ariel DSP32C). The signals were diotically presented via Sennheizer HDA 200 headphones.

The test signal was a 400-msec long 2-kHz tone. It was gated with 200-msec raised-cosine ramps. The test signal was temporally centered in the masker, which consists of a pair of 500-msec long 800-Hz wide frozen-noise maskers, one below and one above the signal frequency. To generate the masker bands, a broadband noise was digitally filtered by setting the magnitude of the Fourier coefficients to zero outside the desired passbands (FFT length = 16000 samples, sampling frequency = 32 kHz). Masked thresholds were measured as a function of the distance Δf from the upper edge of the lower bandpass noise to the signal frequency. The distance Δf was 0, 100, 200, 400, 600, and 800 Hz. For each position of the lower noiseband, four distances $\Delta f + \Delta f_s$ of the lower edge of the upper band to the signal frequency were measured and simulated, with $\Delta f_s = 0$, 200 Hz, 400 Hz and infinity (i.e., only the lower noiseband was presented), respectively. The overall level of each band was 60 dB SPL. The masker was gated with 50-msec raised-cosine ramps.

Three male subjects participated in the experiments, one being the first author JV. All subjects had normal hearing (absolute threshold in quiet $\leq$ 15 dB HL) and no previous history of any hearing problems.

4.4.2 Results

Figure 4.3 shows measured (open symbols) and simulated data (filled symbols) for the experiment B. The level of the test signal at threshold is shown as a function of Δf as a fraction of the signal frequency. The topmost line in each plot shows the data for the symmetric notched-noise condition. The other lines show conditions in which the upper side band is shifted towards higher frequencies. For a certain Δf, thresholds decrease with increasing Δf_s in these conditions. The decrease in threshold relative to the notched-noise condition amounts to up to 10 dB, when only a single lowpass-noise masker is used. The measured data are in agreement with data in the literature (e.g., Patterson and Nimmo-Smith, 1980).

In notched-noise experiments with an asymmetric notch it is assumed that off-frequency channels contribute considerably to signal detection (see introduction). Therefore, a multi-channel version of the model is applied. A modulation-lowpass filter model version was employed, since in experiment A it was shown, that high modulation filters do not contain important information that contributes to signal detection in this kind of experiment. The lower right panel of Fig. 4.3 shows simulated results for the different Δf_s. The general shape of the threshold curves for the different Δf_s notched-noise conditions can be described very well by the model.

To illustrate the strategy of the model, two templates of a 2 kHz signal are shown in Fig. 4.4 for the extreme conditions ($\Delta f_s = 0$, $\Delta f_s = \infty$) of present experiment. In the upper panel of Fig. 4.4, the masker is a notched noise with a symmetric notch ($\Delta f_s = 0$). The distance between the nearest edge of each band and the signal frequency is 400 Hz. In the lower panel, the masker comprises only the lower noiseband ($\Delta f_s = \infty$). Contours are shown on the base of each plot to illustrate the shift of the template in the notched–noise condition (upper panel) towards higher frequencies compared to the template in the lowpass condition (lower panel). This shift may be a visualization of the off-frequency listening strategy discussed in the literature (Patterson, 1976).

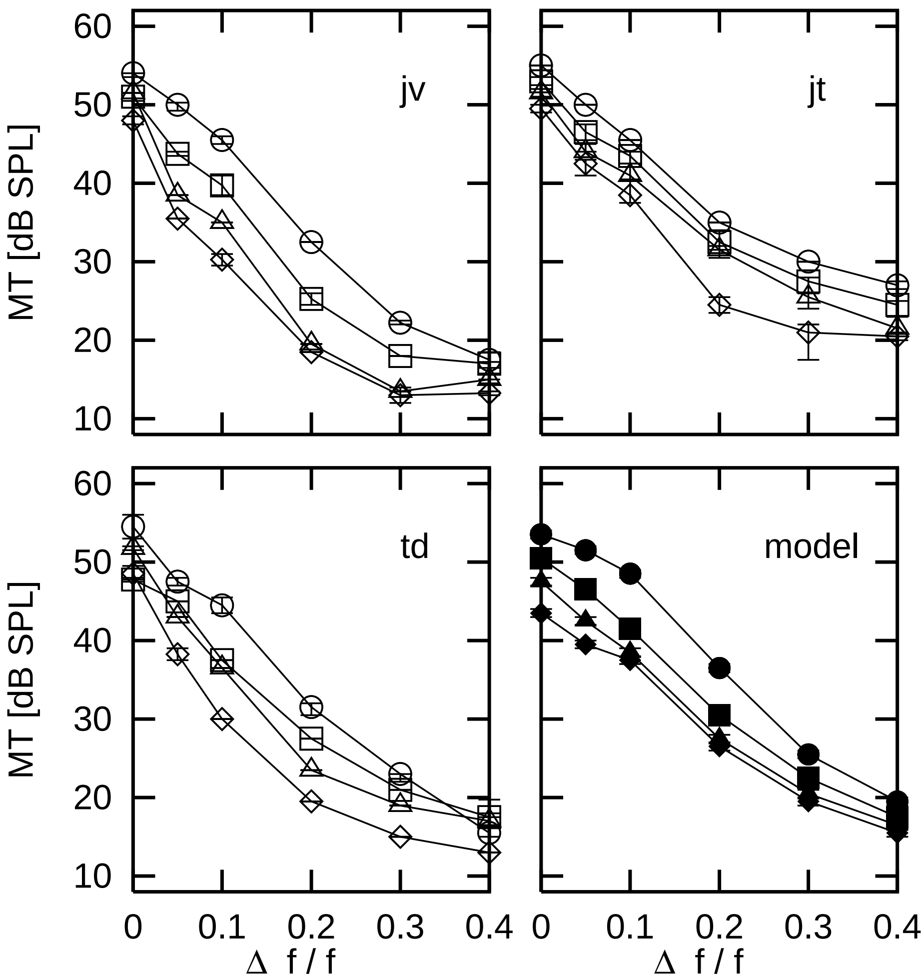

Figure 4.3: *Masked thresholds (MT) for a 2 kHz signal in a notched noise masker as a function of the spectral distance* Δf *from the signal to the upper edge of the lower noise band. The panels with open symbols indicate measured thresholds for three different subjects. Each panel shows data for four conditions: In each condition, a notched noise is used as the masker where the distance of the upper sideband to the signal frequency is* $\Delta f + \Delta f_s$*, with* Δf_s *equal to 0 Hz (circles), 200 Hz (squares), 400 Hz (triangles) and infinity (diamonds).* Δf_s *equal to infinity means, that only the lower noiseband is used. The lower right panel (filled symbols) shows simulated thresholds.*

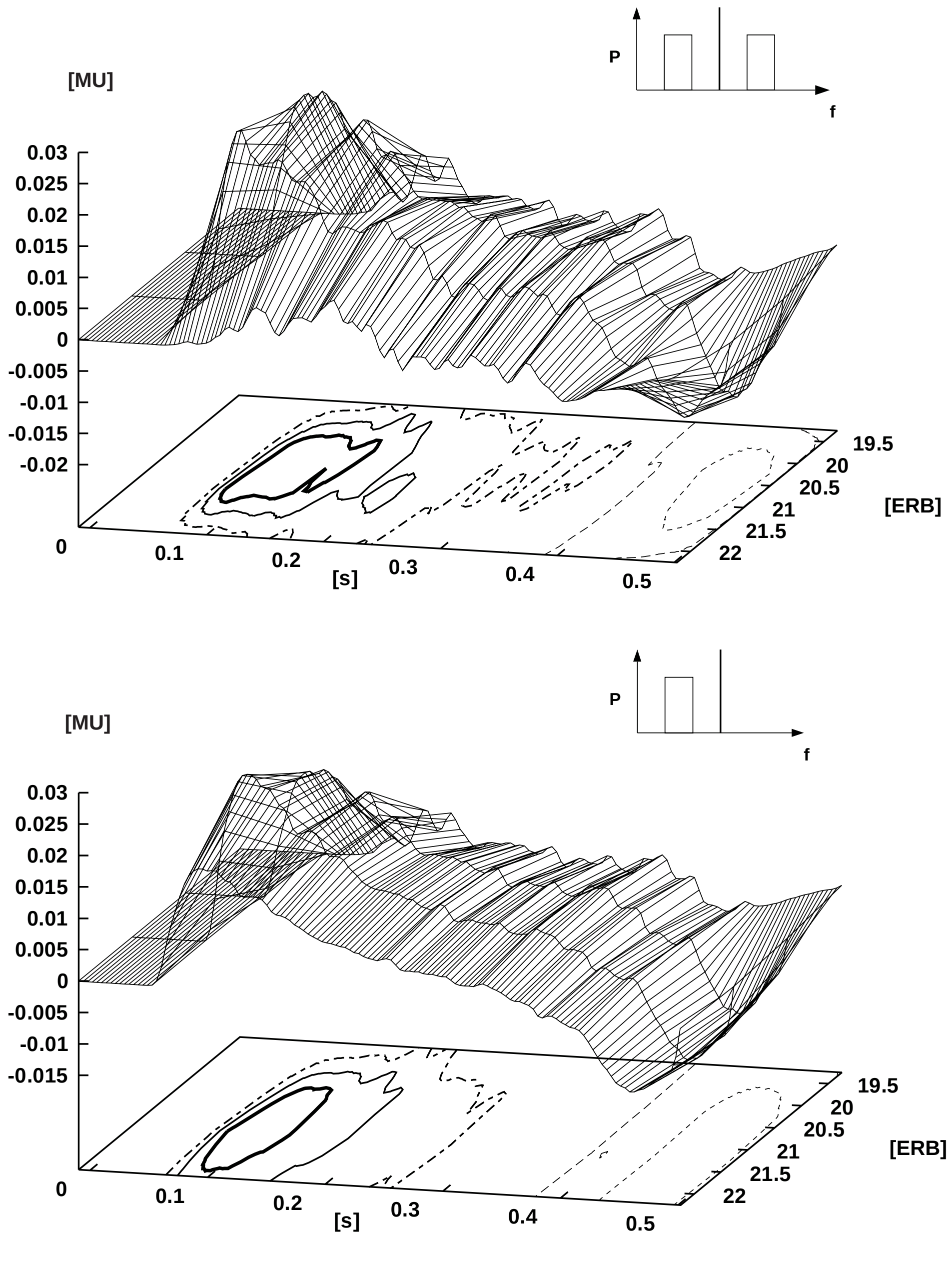

Figure 4.4: *Two templates for a notched-noise condition (upper panel) and for a lowpass-noise condition (lower panel). For details see text.*

4.5 EXPERIMENT C: Asymmetry of masking

4.5.1 Methods

Signals were presented diotically via Sennheizer HD25 headphones. Except for this kind of signal presentation, apparatus and procedure were the same as in experiment A. Bandpass-noise signals with a flat spectrum arithmetically centered at 2 kHz were used as masker and test-signal. To generate the signals, a broadband noise was digitally filtered as in experiment B. The bandwidth of test signal and masker was 4, 16, 64, and 256 Hz, respectively. All combinations of masker and test-signal bandwidth were tested. The masker duration was 700 ms. The test signal had a duration of 500 ms and was temporally centered in the masker, i.e., signal onset was 100 ms after masker onset. 100 ms cosine ramps were applied to the masker and the test signal. Both, test signal and masker were randomly cut out from a fixed noise waveform which was generated separately at the beginning of each experiment. The noise waveform had a periodicity of 4 sec (FFT-length 128k, Sampling frequency 32 kHz). The overall level of the fixed noise waveform was 70 dB SPL.

The level of the test signal at threshold is given in terms of the attenuation level of the 4-sec long noise waveform. Because the noise waveform had the same level of 70 dB for both test signal and masker, this is equal to the signal-to-noise ratio I_T/I_M, where I_T is the intensity of the test signal and I_M is the masker intensity. Note, that the level of the 4-sec long masker waveform was fixed while the level of each realization of the 700-msec long masker undergoes some fluctuations. In fact, for the smallest bandwidth of 4 Hz, the level of the different masker realizations can differ by as much as 5 dB. For 16 Hz, the level variations between the different realization were approximately the same as for 4 Hz. For 64 Hz they amount up to about 3 dB. In contrast, the level of the different realizations do not differ markedly (< 2 dB) for a bandwidth of 256 Hz. Approximately the same level variations occur for the test-signal realizations. The level variation is caused by the slow envelope fluctuations of narrowband-noise signals (cf., e.g., Bos and de Boer, 1966, de Boer, 1966, Hall, 1997, Fig. 4.1)[2]. Thus, the thresholds are measured in a quasi intensity-roving condition for small masker bandwidths, whereas for the 256 Hz condition, the masker has a nearly constant level.

Four subjects (3 male, 1 female) participated in the experiment. One of them was the author JV. As already mentioned in experiment A, the author had extensive prior experience in performing spectral masking experiments. The other three subjects (JD, MT, OW) were inexperienced listeners. The subjects ranged in age from 24 to 29. One subject (MT) received a payment for the participation in the experiments. All subjects had normal hearing (absolute threshold in quiet $\leq$ 15 dB HL) and no previous history

[2] An analytic expression of the standard deviation of the intensity for realizations of bandpass noise with a finite duration was given by de Boer (1966). He showed that the standard deviation σ of the intensity relative to the mean intensity I is equal to $1/(\Delta f * T)^{0.5}$, where Δf represents the bandwidth and T represents the duration of the signal. In the present experiment the calculated standard deviations relative to the mean for the different realizations of the masker ranges from about 0.07 ($\Delta f = 256$ Hz, $T = 700$ ms) to about 0.6 ($\Delta f = 4$ Hz, $T = 700$ ms). Thus, the difference between $I + \sigma$ and $I - \sigma$ ranges from about less than 1 dB to 6 dB.

of any hearing problems.

4.5.2 Results

Figure 4.5 shows individual data and intraindividual standard deviations for 4 subjects. In addition, Fig. 4.6 shows the mean data across subjects and interindividual standard deviations. The figures show the level of the test signal at threshold relative to the masker level as a function of the masker bandwidth. Parameter is the test-signal bandwidth: 4 Hz (circles), 16 Hz (triangles), 64 Hz (squares), and 256 Hz (diamonds).

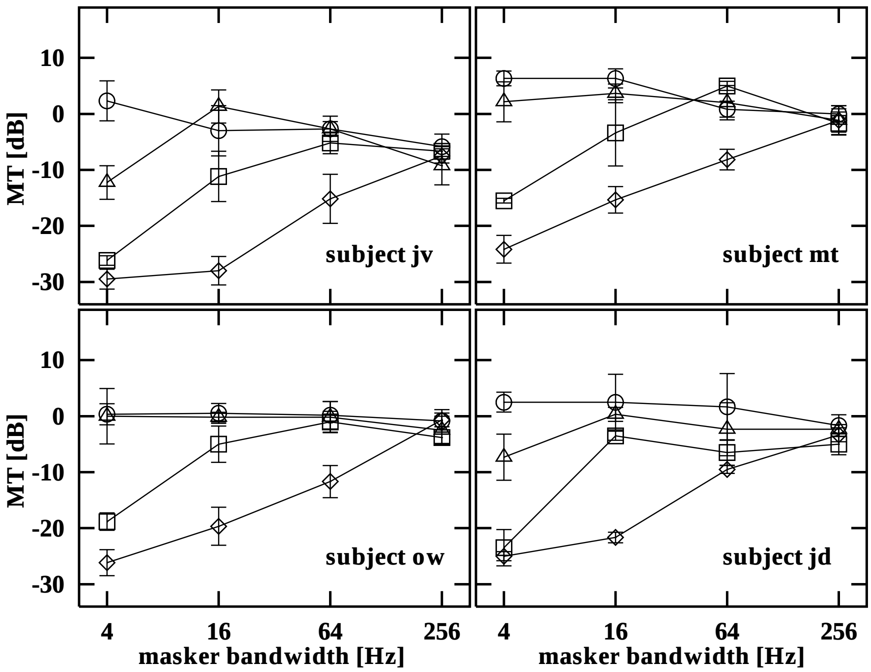

Figure 4.5: *Masked threshold (MT) of a bandpass-noise signal masked by a bandpass-noise masker as a function of masker bandwidth. Parameter is the test-signal bandwidth: 4 Hz (○), 16 Hz (△), 64 Hz (□) and 256 Hz (◇). Individual data are shown for four normal hearing subjects.*

For for three of four subjects, thresholds for a test-signal bandwidth of 4 Hz slightly decrease with increasing masker bandwidth (subject JV, MT, JD). For one subject (OW), thresholds are independent of the masker bandwidth. On average, the threshold for a test-signal bandwidth of 4 Hz is 3 dB for a masker bandwidth of 4 Hz and -2 dB for a masker bandwidth of 256 Hz. The corresponding intensity increment, I_{T+M}/I_M, is approximately 4.7 dB for the masker bandwidth of 4 Hz and 2.1 dB for a masker bandwidth of 256 Hz[3]. The slight decrease of thresholds with increasing masker bandwidth is a consequence of the decreasing level variation with increasing bandwidth (as mentioned above). This was also earlier reported in the literature for tonal signals masked by bandpass noise with different bandwidth (e.g. Bos and de Boer, 1966).

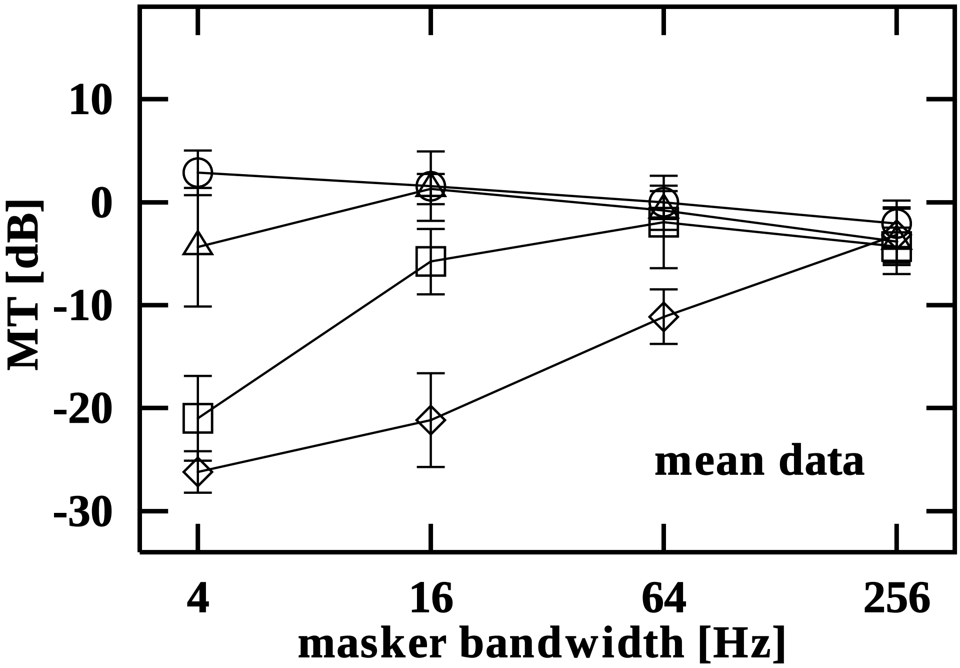

Figure 4.6: *Mean measured data over four subjects for a bandpass-noise signal masked by a bandpass-noise masker as a function of the masker bandwidth. Symbols are the same as in Fig. 4.5.*

Note that the difference limens for bandpass noise, i.e., thresholds in conditions where test signal and masker coincide in bandwidth, do not differ markedly (≤ 1 dB) from the thresholds for the 4-Hz-wide bandpass-noise test signal at different masker bandwidths. This is in agreement with findings by Bos and de Boer (1966): They found a most remarkable agreement between their masking data (a tone masked by a bandpass noise) and their discrimination data (difference limen for bandpass noise) for bandwidths smaller than the critical bandwidth.

An asymmetry between the masking of 4 Hz wide bandpass noise by a 256 Hz wide bandpass noise and the masking of a 256 Hz wide bandpass noise by a 4 Hz wide bandpass noise can be observed for all subjects. On average, the threshold of the 4 Hz wide

[3]Because the signals are not correlated, the intensity of the masker with test signal I_{M+T} can be calculated by adding the intensity of the test signal alone, I_T, and the masker intensity I_M.

bandpass noise is at a level of -2 dB whereas the threshold of the 256 Hz wide bandpass noise is at a level of -26 dB. For a test-signal bandwidth of 256 Hz, thresholds increase monotonically with increasing masker bandwidth. For a masker bandwidth of 256 Hz there is no threshold difference between a test-signal bandwidth of 4 Hz and 256 Hz.

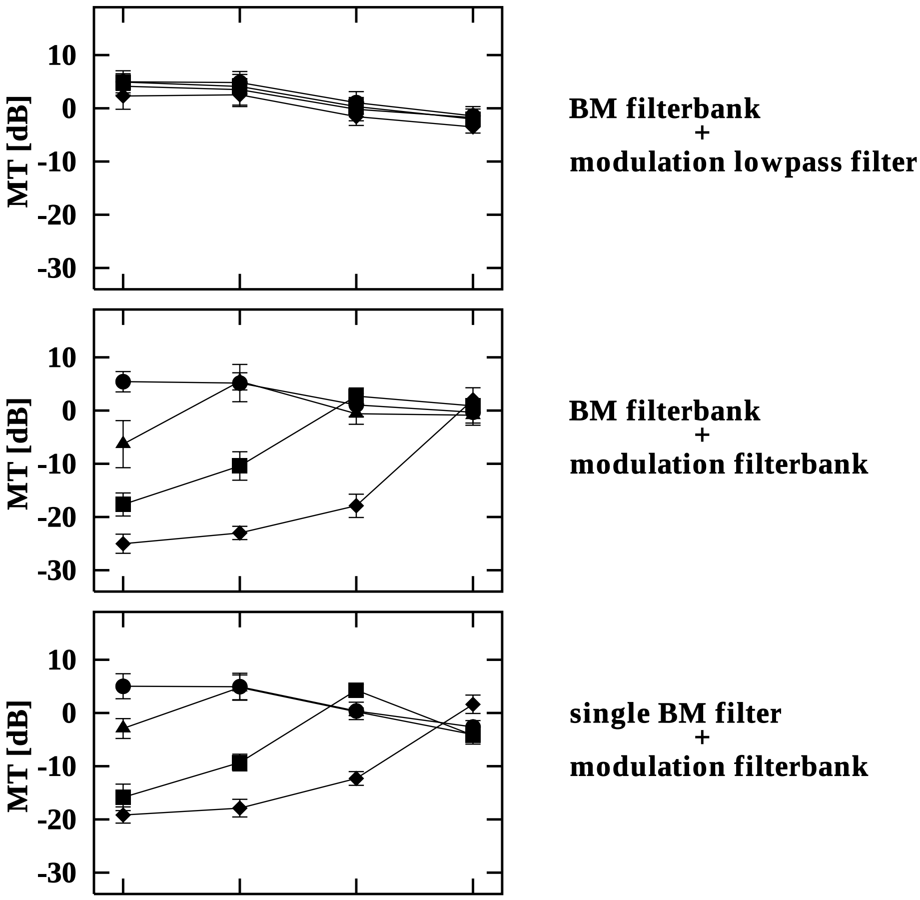

Figure 4.7: *Simulated data for bandpass-noise signal masked by a bandpass-noise masker as a function of the masker bandwidth. Symbols are the same as in Fig. 4.5. Three model-versions are tested: Multi-(peripheral-)channel model with a modulation-lowpass filter (upper panel), multi-(peripheral-)channel model with a modulation filterbank, i.e., the complete model (middle panel), and a single-channel model, where the envelope of the peripheral channel tuned to the center frequency of the signals is analyzed by a modulation filterbank (lower panel).*

In general, for all test-signal bandwidths larger than 4 Hz, two observation can be made: (i) For test-signal bandwidths smaller than the masker bandwidth, thresholds does not depend on the test-signal bandwidth (ii) For test-signal bandwidths larger than the masker bandwidth, threshold decreases with increasing test-signal bandwidth[4]. Thus, in the present experiment, the lowest threshold values are obtained for the largest test-signal bandwidth of 256 Hz.

Figure 4.7 shows simulated thresholds for a test-signal bandwidth of 4 Hz (circles), 16 Hz (triangles), 64 Hz (squares), and 256 Hz (diamonds), respectively. The upper panel of Fig. 4.7 shows predicted thresholds obtained with the same model as used in experiment B, where only a lowpass-filtered envelope is processed in each peripheral filter. The model accounts for the decreasing thresholds with increasing masker bandwidth for the 4-Hz-wide bandpass-noise signal. In addition, in agreement with the measured data, difference limens for bandpass noises and thresholds for the 4-Hz-wide bandpass-noise signal at different masker bandwidths do not differ markedly. However, the simulated threshold curves are nearly independent of the test-signal bandwidth. This shows that a model, which is mainly based on the energy of the output of the different peripheral filters, is inadequate to account for the effect of asymmetry of masking.

The middle panel of Fig. 4.7 shows simulated thresholds with the complete model, where the envelope is analyzed by a modulation filterbank in each peripheral channel. There is a good agreement between simulated data and measured data. For a test-signal bandwidth of 4 Hz, the predicted thresholds for the masker bandwidth of 256 Hz and for 4 Hz differ by about 5 dB, which agrees well with the experimental data. Furthermore, for test-signal bandwidths larger than the masker bandwidth, simulated thresholds decrease with increasing test-signal bandwidth. This is also seen in the measured data. The difference between the threshold of a 4 Hz wide bandpass noise masked by a 256 Hz bandpass noise and the threshold in the reversed condition can be predicted quantitatively by the model. The predicted threshold of a 4 Hz wide test signal masked by a 256 Hz wide masker is -0.3 dB whereas it is -26 dB in the reversed condition. Both values coincide well with the average measured thresholds of -2 dB and -26 dB, respectively (see above).

The role of off-frequency information is investigated using a single-channel version of the model, where the information from the single peripheral filter tuned to the center frequency of the signals is processed. Thresholds obtained with this model are shown in the lower panel of Fig. 4.7. There is a qualitative agreement between simulated and measured data. The decrease in threshold with increasing test-signal bandwidth for a fixed masker bandwidth is slightly smaller than that obtained with the multi-channel version of the model (middle panel of Fig. 4.7). For example, the simulated threshold for the 256 Hz wide test-signal bandwidth masked by a 4 Hz wide bandpass noise is only -20 dB, in contrast to -26 dB obtained with the “complete” model. It is interesting to note, however, that most of the asymmetry of masking can be already predicted by such

[4]For a test-signal bandwidth of 16 Hz, two subjects (OW, MT)) did not show a difference between threshold curves of 4 Hz and 16 Hz. However, on average, the threshold of a 16 Hz bandpass noise masked by a 4 Hz bandpass noise is lower than for a 4 Hz bandpass noise masked by 4 Hz bandpass noise.

a single-channel model.

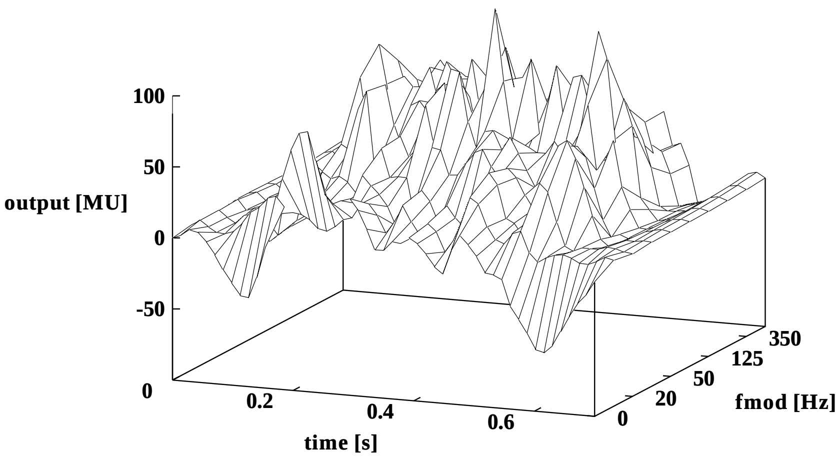

Figure 4.8: *Internal representation of the test signal calculated as the difference between the actual internal representation of the test signal plus masker and the stored mean internal representation of the masker alone. The upper panel shows the internal representation of a test signal with a bandwidth of 4 Hz, in the lower panel the test-signal bandwidth was 256 Hz. In both conditions the masker bandwidth was 4 Hz and the test-signal level was 15 dB above the masker level of 70 dB SPL.*

To illustrate the detection cues in the different masking conditions in the framework of the model, the difference between the internal representation of signal plus masker and that of the masker alone is shown in Fig. 4.8 for two extreme conditions. In the upper panel of the figure, the test signal and the masker have the same bandwidth of

4 Hz. In the lower panel, the test-signal bandwidth is 256 Hz and the masker bandwidth is 4 Hz. In both conditions, the level of the test signal was 15 dB above the level of the masker, i.e., in both conditions the test signal was presented well above threshold. The figure shows the internal "activity" as a function of time and center frequency of the modulation filters (cf., 6 in Dau et al., 1997a). The ordinate is scaled in model units (MU).

The two activity patterns differ markedly. For a 4 Hz wide test signal, the test signal mainly activates the modulation-lowpass filter. In addition, the onset (100 ms after masker onset) and offset (600 ms after masker onset) of the test signal are also represented in the modulation filters centered at higher modulation frequencies. In contrast, the activity pattern for the 256 Hz wide test signal is dominated by high modulation frequencies. This indicates that in conditions where the test-signal bandwidth is larger than the masker bandwidth probably higher modulation frequencies are used by the subject as an additional cue. For Gaussian bandpass noise (as used in the present experiment) the envelope of the signal will fluctuate with rates from 0 to the bandwidth of the noise (cf., Lawson and Uhlenbeck, 1950, Dau et al., 1997a). Thus, in conditions, where the test-signal bandwidth is larger than the masker bandwidth, the modulation spectrum of the test signal will be broader than that of the masker. In these conditions, the higher (unmasked) modulation frequencies of the test signal will offer an additional detection cue. In contrast, in the reversed condition, when the test-signal bandwidth is smaller than the masker bandwidth, the detection is mainly based on the energy within the critical band filter, since the inherent envelope fluctuations of the test signal are completely masked by the inherent envelope fluctuations of the masker.

4.6 DISCUSSION

The model accounts (i) for spectral masking data derived from a typical symmetrical notched-noise condition, (ii) for the notched-noise masking data in typical off-frequency conditions, and, (iii) for the effect referred to as "asymmetry of masking", where thresholds are much lower in conditions with test-signal bandwidths larger than the masker bandwidth as compared to the reversed condition (, i.e., where the masker bandwidth is larger than the test-signal bandwidth).

4.6.1 Notched-noise masking conditions

The present model contains the gammatone filterbank as the peripheral filter stage, which has been initially developed by Patterson et al. (1987) to account for the masking data with notched-noise maskers. Thus, the present work may not provide substantial new information. Note, however, that the present model differs in the detection strategy from an energy-detection model as, for example, used by Patterson (1976) and Patterson and Nimmo-Smith (1980) to derive auditory filter shapes. First, while the energy-detector model uses a signal-to-noise-ratio calculation, the decision strategy of the present model is based on a matched-filtering process of the actual activity pattern in a trial in comparison with an internal representation of the signal to be detected (the template). Second, Patterson (1976) and Patterson and Nimmo-Smith (1980) determined a specific filter position yielding the largest signal-to-noise ratio to account for the data in typical conditions of off-frequency listening. In contrast, the present model processes the internal representation of the output of several gammatone filters and combines them optimally, assuming independent observations at several channel outputs. The advantage of the present approach is that the weighting of the relevant filters occurs automatically (by the template) and the model allows spectral integration of information.

4.6.2 Difference limen for narrow-band noise

Several authors have reported a decreasing intensity difference limen for bandpass-noise signals with increasing bandwidth, for noise bandwidths smaller than the critical bandwidth (e.g., Zwicker, 1956, Zwicker and Fastl, 1990, Bos and de Boer, 1966). A decreasing difference limen for bandpass-noise signals (, i.e., threshold in the condition where the test-signal and the masker coincides in bandwidth,) with increasing bandwidths is also observed in our data. On average, the intensity difference limen for the 4-Hz-wide bandpass noise is 4.7 dB, whereas it is only 2.1 dB for the 256-Hz wide bandpass noise. A power-spectrum model would predict a constant threshold independent of the noise bandwidth. Zwicker (1956) and Bos and de Boer (1962) suggested that the statistical properties of the envelope fluctuations of narrow-band noise explain the increased thresholds. De Boer (1966) showed that a model that analyzes the energy distribution, i.e., the mean and standard deviation of the energy, can account for intensity discrimination of fluctuating signals. According to de Boer, the increasing threshold with decreasing bandwidth is a consequence of the increasing standard deviation relative to the mean with decreasing bandwidth.

In Fig. 4.7, it was shown that the present model accounts for the increasing difference limen with decreasing noise bandwidth. Interestingly, the model predictions based only on the lowpass-filter output of the envelope agree well with the data (upper panel of Fig. 4.7), indicating that processing of modulations only play a minor role in these experiments. Thus, as in the analytical calculations by de Boer (1966), the increasing variance of the energy with decreasing bandwidth is the reason for the increasing threshold. However, in the present model the variance of the signal energy is not calculated explictly as in de Boer. Instead, the present model simulates the course of the adaptive experiment explicitly, so that the statistics of the stimuli at the output of the preprocessing stages determine threshold.

4.6.3 Role of modulations in asymmetry of masking conditions

Whereas for the notched-noise detection experiments and for the intensity-discrimination experiments with narrow-band noise, the processing of the energy in each gammatone filter was sufficient to account for the experimental data, the conditions of asymmetry of masking can only be described with information from modulation filters tuned to higher modulation rates (cf., Fig. 4.7). Most of the effect can already be predicted with an analysis performed in only one peripheral filter, tuned to the center frequency of the noise band. The additional detection cue in conditions, where the modulation spectrum of the test signal is broader than that of the masker, was illustrated in the lower part of Fig. 4.8. Because the model assumes frequency selectivity for modulations, the model is able to use this difference in the modulation spectrum as an additional cue. Even a modulation-lowpass approach with a higher cut-off frequency, as proposed by Viemeister (1979), would not be able to describe asymmetry-of-masking effects, since such an approach can not separate the higher inherent test-signal fluctuations from the inherent masker fluctuations.

The predictions with the multi-channel model showed the best agreement with the experimental data. This indicates that off-frequency filters contribute additional information in the conditions where the test-signal bandwidth is larger than the masker bandwidth. As expected, the difference between the thresholds obtained with the single-channel model and the multi-channel version increases with increasing test-signal bandwidth. The largest difference (6 dB) was found for a test-signal bandwidth of 256 Hz and a masker bandwidth of 4 Hz. However, an off-frequency listening strategy as proposed in the case of the asymmetric notched-noise conditions from Experiment B, is unlikely in the present experiment, since the signal bandwidth is always smaller than the critical bandwidth. It is not the energy from off-frequency channels but the modulations from off-frequency channels, which are used to improve signal detection in conditions where the test-signal bandwidth is larger than the masker bandwidth. Indeed, for a masker bandwidth of 4 Hz, the predicted threshold for a 256 Hz test-signal bandwidth is only 2.5 dB lower than for test-signal bandwidth 4 Hz, when the envelope in each peripheral filter is processed by a modulation-lowpass filter instead of a modulation filterbank (upper panel of Fig. 4.7).

In the present study, only stochastic signals (bandpass noises) were used for both test signal and masker. In typical conditions denoted as asymmetry of masking in the literature, the masking of a tone by a narrow-band noise was compared to the reversed condition, i.e., masking of a narrowband noise by a tone (e.g., Hellman, 1972, Greenwood, 1961, 1971, Moore et al., 1998). Thus, in these cases, the masking properties of a deterministic signal are compared with the masking properties of a stochastic signal. However, the asymmetry of masking observed between a tone and a bandpass noise may be interpreted in the same way as described in Experiment C of the present study. The inherent envelope fluctuations of the bandpass-noise test signal can be used as an important detection cue, when masked by a tone which has no inherent envelope fluctuations. In the reversed condition, the analysis of modulations would not improve the detectability of the signal, since only the masker excites the modulation filters. In a recent paper by Moore al. (1998), a similar explanation was proposed. They argued that the lower thresholds in the conditions with a tonal masker and a noise signal compared to the reversed condition may be attributed to the availability of a within-channel cue of a fluctuation in level. However, although they discussed the possible envelope fluctuations, they did not model the effect explicitly. The ability of the modulation-filterbank model to describe the different masking properties of a bandpass-noise and a tone are subject to recent investigations (Derleth et al., 1998).

4.7 Conclusions

- The present study examined the ability of the modulation-filterbank model presented by Dau et al. (1997a+b) to account for spectral-masking data. It is shown that the model accounts for frequency selectivity in notched-noise experiments. This mainly is a consequence of the gammatone-filter stage of the model.

- Spectral masking in off-frequency listening conditions can be accounted for, which results from the combination of information across critical bands as realized in the optimal decision stage.

- The model also accounts for the increase of intensity-discrimination thresholds (difference limen) for narrow-band noise with decreasing noise bandwidth in sub-critical bandwidth conditions, where the bandwidth is smaller than the critical band. This results from the simulated experimental algorithm, which ensures that the model's resolution is limited by the external statistics of the signal.

- In contrast to the energy-detector model, the present model is also sensitive for temporal effects of spectral masking which plays a role in conditions of asymmetry of masking. This effect can be predicted by the model, since also changes in the modulation spectrum are analyzed apart from the energy difference, when the signal is added to the masker. This is particularly effective in conditions, where the masker bandwidth is smaller than the test-signal bandwidth: the model uses the information from modulation filters tuned to high modulation frequencies, which are mainly excited by the inherent fluctuations of the test signal. A model which only processes the energy of the signal fails in these conditions. Although the role of temporal processing was discussed qualitatively in the literature, to our knowledge the present study is the first to show that a model with an appropriate processing of the time structure of the signals is able to quantitatively predict the effect of asymmetry of masking.

Chapter 5

Within-channel cues in comodulation masking release (CMR): Experiments and model predictions using a modulation-filterbank model[1]

ABSTRACT

Experiments and model calculations were performed to study the influence of within-channel cues versus across-channel cues in comodulation masking release (CMR). We considered a class of CMR experiments that are characterized by a single (unmodulated or modulated) bandpass noise masker with variable bandwidth centered at the signal frequency. A modulation-filterbank model suggested by Dau et al. [J. Acoust. Soc. Am. **102**, 2892-2905 (1997)] was employed to quantitatively predict the experimental data. Effects of varying masker bandwidth, center frequency, modulator bandwidth, modulator type, and signal duration on CMR were examined. In addition, the effect of bandlimiting the noise before or after modulation was shown to influence the CMR in the same way as a systematic variation of the modulation depth. It is demonstrated that a single-channel analysis, which analyzes only the information from one peripheral channel, quantitatively accounts for the CMR in most cases, indicating that an across-channel process is generally not necessary for simulating results from this kind of experiment.

[1]This chapter is a slightly modified version of Verhey, J.L., Dau, T., and Kollmeier, B. (1998) "Within-channel cues in comodulation masking release (CMR): Experiments and model predictions using a modulation-filterbank model", submitted to JASA.

5.1 INTRODUCTION

Detection thresholds for a test signal presented in a noise masker can be markedly lower when the masker is modulated compared to the reference condition where the masker is unmodulated. This was first shown by Hall et al. (1984). In one of their experiments they measured the detectability of a sinusoidal signal in the presence of a bandpass-noise masker which was spectrally centered at the signal frequency. They found that for masker bandwidths larger than the critical band signal thresholds in the modulated-noise condition were up to 9 dB lower than in the reference condition where unmodulated bandpass noise was used as the masker. This difference between the modulated and unmodulated condition diminishes when the masker bandwidth is smaller than the critical bandwidth. Based on these findings, Hall et al. suggested that the observers may compare the envelope fluctuations between different auditory channels, e.g. they took advantage of the correlated masker envelope across frequency bands in the modulated condition. Hence they called the difference in threshold between unmodulated and modulated condition "comodulation masking release" (CMR).

In the literature the CMR experiments can be divided in two different classes due to different masker types. In the first class of experiments the masker is a single bandpass noise centered at the signal frequency (as just described) (Hall et al. 1984, Haggard et al. 1990, Schooneveldt and Moore, 1989a, Carlyon et al., 1989, Hall et al. 1996). In the second class of experiments the masker consists of several masker bands, one at the signal frequency (on-frequency band) and one or more other bands (flanking bands) spectrally separated from the on-frequency band. It has been shown that adding one comodulated band to an on-frequency band can improve the detectability of the sinusoidal signal by up to 8 dB (e.g. Hall et al., 1984, Hall et al., 1989, Schooneveldt and Moore, 1987, 1989b, Fantini et al., 1993, Eddins and Wright, 1994, Hatch et al., 1995). In contrast to the experiments of the first class, the difference between the reference condition (on-frequency band alone) and the comodulated condition is largest for a small spectral distance between the two masker bands (smaller than the critical bandwidth) and decreases with increasing spectral distance. It was argued, that some of the effect in the second class of experiments might result from beating between components of the two masker bands within one peripheral channel, i.e., it results from a within-channel cue (McFadden, 1986, Schooneveldt and Moore, 1987). For large spectral distances, however, the threshold difference between reference and comodulated condition (e.g. Moore, 1992) amounts to about 2-6 dB. It is a assumed that this is a consequence of an across-channel process and is therefore called "true" CMR.

Several authors also discussed the most appropriate reference condition to define CMR in the second class of experiments. CMR is normally obtained by subtracting the thresholds with no flanking band from thresholds with comodulated flanking band(s). However, in some studies another definition of CMR is proposed: CMR is defined as the difference between thresholds with comodulated flanking bands and thresholds with non-comodulated flanking bands. This latter convention is more similar to the definition of CMR in the first class of experiments. In most cases it gives larger values for CMR because thresholds are generally higher in the presence of non-comodulated flanking bands than for the on-frequency band alone even for large spectral distances (e.g., Schooneveldt and Moore, 1987).

Concerning the first class of experiments, Carlyon et al. (1989) argued that the difference between the unmodulated and modulated conditions is not an accurate way to determine "true" CMR, because in some cases a modulation of the masker produce a masking release even for subcritical masker bandwidths, where the role of across-channel processes is unclear (Carlyon et al., 1989; Schooneveldt and Moore, 1989a). Hence Carlyon et al. (1989) suggested subtracting the threshold difference for a subcritical condition from the threshold difference obtained for bandwidths larger than the critical bandwidth to get a measure of an "across-channel" CMR. This would be more or less equivalent to the definition of CMR in the second class of experiments where the on-frequency band-alone condition was proposed as the reference. However, as mentioned by Hall et al. (1995), this definition of CMR is somewhat problematic, because it assumes that there is a clear distinction between bandwidths where "pure" within-channel processes are the reason for the threshold difference and bandwidths where across-channel processes occur as well. Furthermore, the above definition depends on the definition of the critical bandwidth.

Schooneveldt and Moore (1989a) explained the amount of CMR in the first class of experiments in terms of envelope statistics within one critical band (within-channel process). They argued that changes in the statistical properties of the envelope produced by adding the signal to the noise can produce threshold reduction also for noise bandwidths smaller than the critical band. Because of this within-channel cue, which facilitate the detection of the signal in modulated noise, they stated that the magnitude of the "true" CMR is overestimated, when it is defined as the difference in thresholds between modulated and unmodulated condition.

Although several studies have qualitatively discussed the role of within-channel processes in CMR experiments, there is no study that quantifies the amount of masking release due to within-channel cues. Therefore, simulations with a (single-channel) model suggested by Dau et al. (1997a) were performed here and compared with experimental data. The model analyzes the envelope of the peripheral channel tuned to the signal frequency by a modulation filterbank. The model of Dau et al. has been previously tested and evaluated in a variety of modulation detection and discrimination conditions.

This study will focus on the first class of experiments. The experiments were replicated and compared to simulated data. Three additional experiments were performed to analyze the effect of waveform generation of the modulated masker.

5.2 SINGLE-CHANNEL MODEL

In Dau et al. (1997a) a model was proposed to describe modulation detection and modulation masking. The model combines several preprocessing stages with an optimal detector as the decision device (see Fig. 5.1). The first stage of the model is the linear gammatone filterbank model of Patterson et al. (1987), that simulates the bandpass characteristic of the basilar membrane. To investigate within-channel cues in the present study, a single-channel version of the model is applied, i.e., only the output of the gammatone filter centered at the signal frequency was further processed. The signal at the output of the specific filter was halfwave rectified and lowpass filtered at 1000 Hz, i.e., for high center frequencies only the envelope of the signal was further processed. To simulate the adaptive properties of the auditory periphery, a chain of five consecutive nonlinear feedback loops was incorporated (Püschel, 1988). The feedback loops were initially developed to describe forward masking experiments. In addition to adaptation, this stage transforms stationary signals approximately to the logarithm of the input. In contrast, fast fluctuations are transformed nearly in a linearly way. To further analyze the amplitude fluctuations of the envelope, a linear modulation filterbank is incorporated in the model. The frequency selectivity for modulation frequencies is based on physiological findings and psychoacoustical data to modulation masking and modulation detection. The center frequency of the highest modulation filter was set to a quarter of the center frequency of the peripheral filter. This slight modification of the model was motivated by physiological findings by Langner and Schreiner (1988). Apart from that, all parameters of the model were identical to those described by Dau et al. (1996a+b, 1997a+b). To model the limits of resolution, an internal noise with a constant variance was added to the output of each modulation filter output. In the decision process, a stored temporal representation of the signal to be detected (the template) is compared with the actual activity pattern by calculating the cross correlation between the two temporal patterns. This is comparable to a "matched filtering" process. Because in the experiments statistical stimuli were used, thresholds can vary between different simulated runs of the experiment. Therefore, as the final simulated threshold, the mean (and standard deviation) of twenty simulated experimental runs was used.

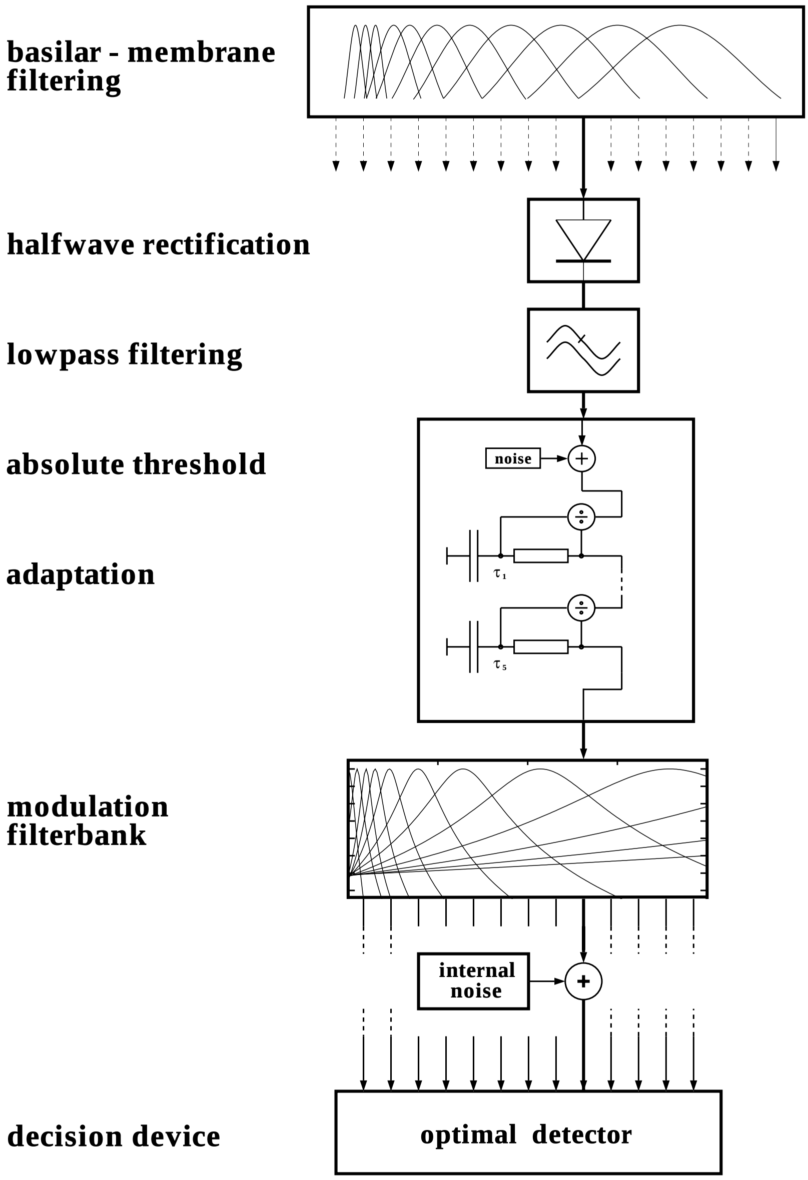

Figure 5.1: *Block diagram of the psychoacoustical model as described in Dau et al. (1997a). The signals are preprocessed, subjected to adaptation, filtered by a modulation filterbank and finally added to internal noise. The decision device is realized as an optimal detector.*

5.3 METHODS

5.3.1 Procedure

In all experimental conditions, a 3-interval forced-choice procedure (3 IFC) was used to measure the masked threshold of a sinusoidal signal in the presence of a noise masker. A trial contained three intervals separated by 500 ms of silence. Two intervals consisted of the masker alone, and one randomly chosen interval contained the masker plus signal. The signal was presented well above threshold at the beginning of the experiment. Subjects had to indicate the interval with the signal by depressing the corresponding key on a keyboard. Feedback was provided on a screen in front of the subjects as to whether the response was correct or not. The level of the signal was varied according to a one-up two-down procedure. This procedure estimates the 70.7 % point on the psychometric function (Levitt, 1971). The step size was 8 dB at the beginning of the experiment. It was reduced to 4 dB after the first reversal and finally to 2 dB after the second reversal. Then the step size was held constant over the next 10 reversals. The threshold was determined by calculating the median of the levels during the last 10 reversals. The final threshold estimate was taken as the mean over three runs.

5.3.2 Apparatus and Stimuli

The stimuli were generated digitally with a sampling rate of 32 kHz. Stimulus generation and presentation were controlled by a silicon graphics workstation (INDY), which also sampled the listener's responses and controlled the procedure. The software package SI was used for signal generation and control of the experiments, which was developed at the Universität Göttingen. The stimuli were D/A converted (16 bits), and then preamplified and lowpass-filtered at 16000 Hz with a computer controlled audiometric amplifier developed in a joint research project on speech audiometry (Kollmeier, 1996). The subjects were situated in a sound-attenuating booth. The stimuli were presented diotically via a Sennheizer HDA 200 headphone without free-field equalization.

The masker was a band-limited noise centered at the signal frequency, i.e., at 1 kHz, 2 kHz and 4 kHz, respectively. The masker bandwidth was 50, 100, 200, 400, 1000 and 2000 Hz. In general, two different noise generation processes were used. In the "reference condition" a digitally band-pass filtered Gaussian noise was applied. In the "comodulated condition" a broad-band noise (from 0 to 10000 Hz) was first multiplied with a digitally filtered low-pass noise and finally restricted to the desired bandwidth. The cut-off frequency of the lowpass noise (modulator bandwidth) differed in the experiments. The spectrum level of the bandpass noise was 30 dB. In the last three experiments, a different "comodulated noise" was used (see below). The masker duration was 600 ms. This period was randomly cut out from a fixed noise waveform which was generated at the beginning of each experiment and had a periodicity of 3 s. 10 ms cosine ramps were applied to the masker.

In all experiments the sinusoidal signal had a duration of 300 ms (except for the third experiment where the effect of varying signal duration was tested explicitly), and was gated with 50 ms cosine ramps if not stated otherwise. The signal onset was 150 ms after the masker onset.

5.3.3 Subjects

Five normal hearing subjects (2 male, 3 female) ranging in age between 24 and 32 years participated voluntarily in the experiments. Two of them were the authors JV and TD. One subject participated in all experimental conditions. At least three subjects were used for each experiment. All subjects had prior experience in psychoacoustic experiments.

5.4 RESULTS

5.4.1 Experiment 1: CMR as a function of carrier frequency and masker bandwidth

In this experiment the dependence of CMR on the center frequency CF of the noise band was investigated. The signal frequency was 1, 2 and 4 kHz. The modulator bandwidth Δf_{mod} was chosen at 50 Hz. The experiment is similar to that described by Haggard et al. (1990). It was replicated here with a slightly different threshold procedure and method of stimulus generation.

Figure 5.2 shows measured individual data for four subjects (upper and middle panels), mean data across subjects (lower left panel) and simulated data (lower right panel) for a signal frequency of $CF = 2$ kHz. The figure shows masked thresholds of a sinusoidal signal as a function of the masker bandwidth. The circles represent thresholds for the reference condition whereas the squares represent thresholds for the comodulated condition. In the reference condition the threshold first increase with increasing bandwidth and then reaches a constant value for masker bandwidths $\Delta f \geq 400$ Hz. With the classical procedure of fitting two lines to the data (Fletcher, 1940), a critical bandwidth of about 150 to 300 Hz would be predicted depending on the subject. The average across subjects yields a critical bandwidth of about 220 Hz which is somewhat smaller than the critical band widths of 240-310 Hz reported in the literature (Scharf, 1970 , Moore and Glasberg, 1983; Schooneveldt and Moore, 1989a). In the comodulated condition, threshold is independent of masker bandwidth for $\Delta f \leq 200$ Hz whereas threshold decreases with increasing masker bandwidth for $\Delta f > 200$ Hz. The difference in threshold between the reference condition and the comodulated condition (CMR) increases with increasing bandwidth. For large bandwidths ($\geq$ 1000 Hz) this CMR value converges at a value of 8 to 14 dB, depending on the subject.

The simulated thresholds (lower right panel) are in good agreement with experimental results. However, there is a constant threshold shift of about 4 dB towards higher thresholds. In the reference condition threshold increases with increasing masker bandwidth up to the critical bandwidth and then remains at a constant level (of 53 dB) whereas in the comodulated condition threshold decreases with increasing bandwidth. The model predicts a CMR effect of 11 dB for large masker bandwidths which is the same as in the experiment. Figure 5.3 shows averaged experimental data of three subjects (open symbols) and simulated data (filled symbols) for three signal frequencies: 1 kHz (upper panels), 2 kHz (middle panels), and 4 kHz (lower panels)[2]. The circles represent thresholds for the reference condition and squares show thresholds for the comodulated condition.

The thresholds for a masker bandwidth of 50 Hz vary very little with center frequency and are more or less the same in both conditions. For all center frequencies the threshold in the reference condition first increases with increasing masker bandwidth up

[2]Subject KS only participated in the experimental condition with CF = 2 kHz. Therefore, to compare the data across signal frequencies, for CF = 2 kHz only the data over subject KW, TD, and JV were averaged.

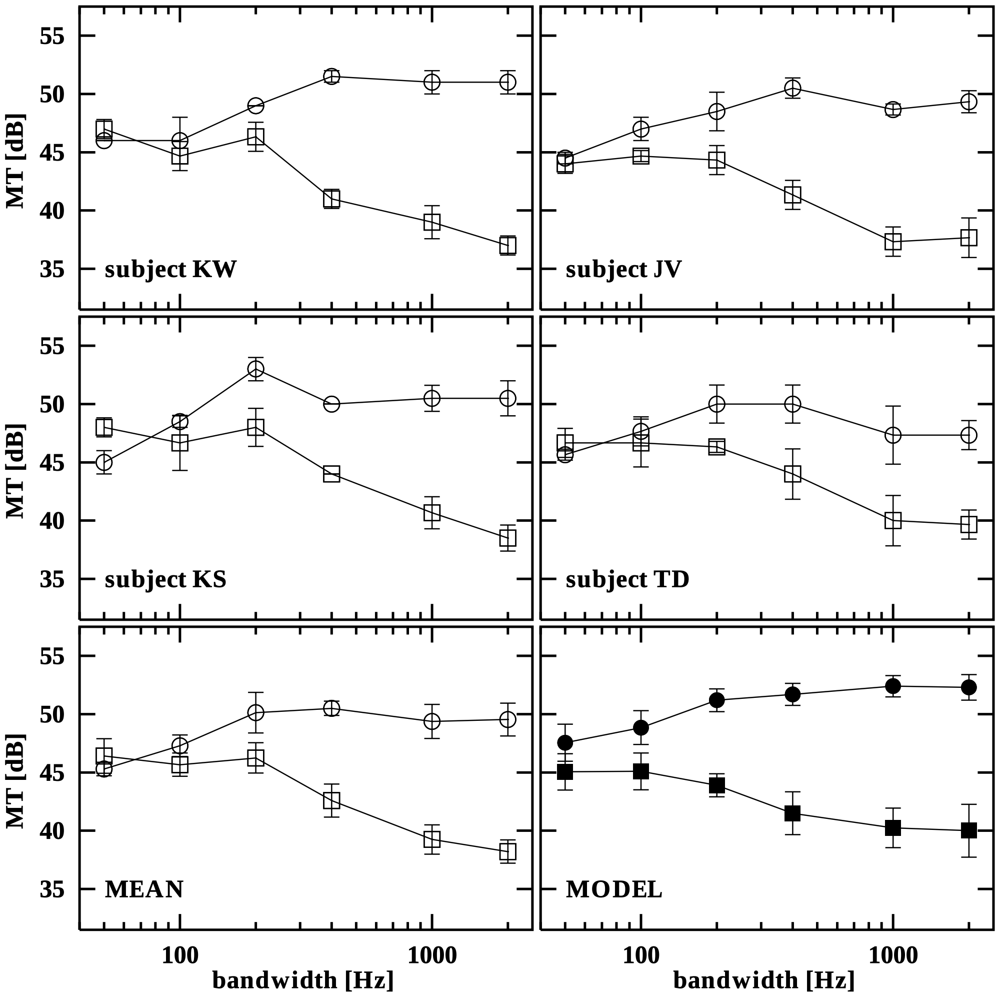

Figure 5.2: *Signal threshold as a function of masker bandwidth in random noise (○) and modulated noise (□). The modulator bandwidth was 50 Hz, the signal frequency was 2 kHz. The upper four panels show individual data. The lower left panel shows the average data across all subjects, the lower right panel shows simulated data.*

to the maximum threshold at about 49 - 53 dB depending on center frequency. In the "comodulated condition" the thresholds remain constant or slightly increase up to $\Delta f =$ 100 Hz for $CF = 1$ kHz and up to $\Delta f = 200$ Hz for $CF = 2$ kHz and 4 kHz. Beyond this masker bandwidth thresholds decrease with increasing masker bandwidth. The maximum threshold difference between reference and comodulated condition (CMR) for large

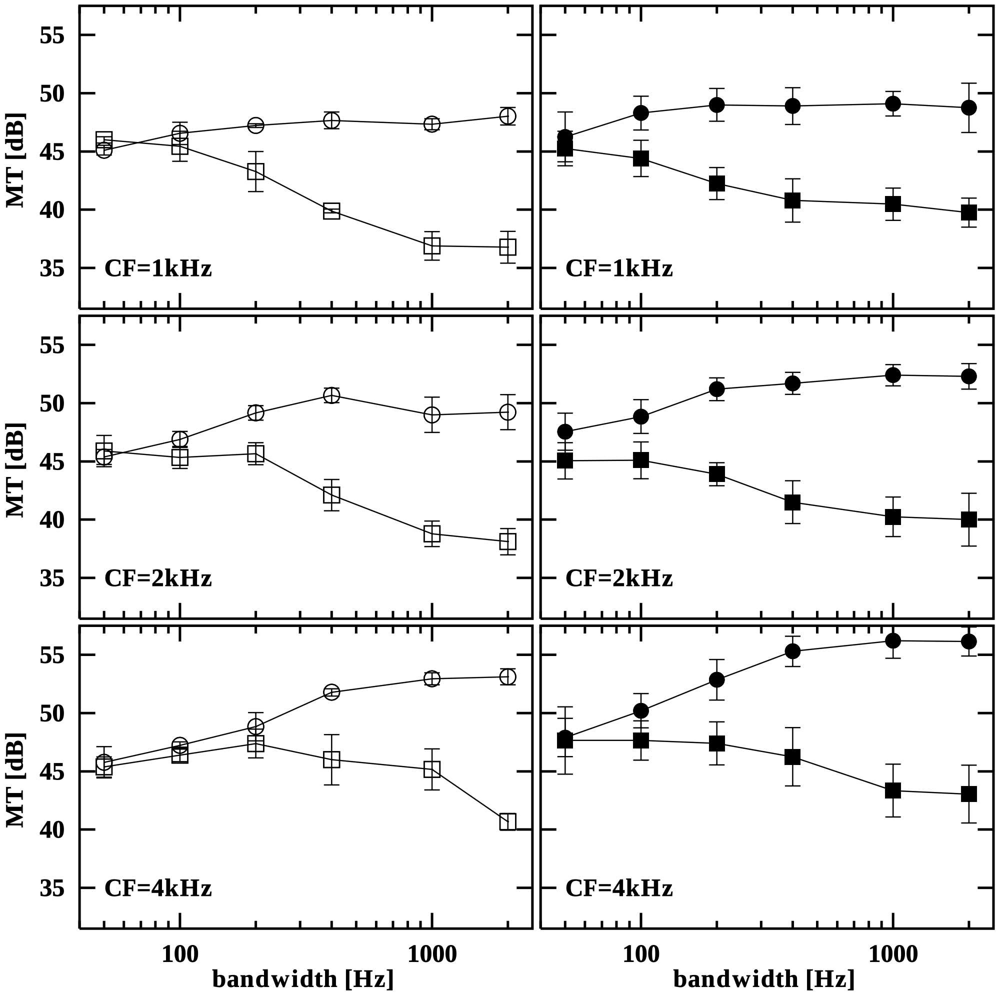

Figure 5.3: *Mean measured data over three subjects (left row of panels) and simulated thresholds (right row of panels) for a signal presented in random noise (○) and modulated noise (modulator bandwidth of 50 Hz, □) as a function of masker bandwidth. The signal frequency was 1 kHz (upper panels), 2 kHz (middle panels), and 4 kHz (lower panels).*

masker bandwidths ($\geq$ 1000 Hz) is 11 dB for $CF = 1$ and 2 kHz and 13 dB for $CF =$ 4 kHz. The thresholds are 10 dB lower than in a comparable experiment by Hall et al. (1984). This shift is due to a different spectrum level in the two studies (40 dB in their experiment versus 30 dB in the present study). In addition, the general shape of our data deviate slightly from the experimental results presented by Haggard et al. (1990). They found, that the "CMR existence region" (all masker bandwidths with CMR greater than

+3dB) varies with center frequency in a way similar to the critical bandwidth. Our data do not show such a direct correlate. There is a negligible effect of center frequency on the "CMR existence region". This difference might be due to a different way of generating the noise masker. Haggard et al. generated bandpass noise by multiplying lowpass noise with a sinusoid at the center frequency of the desired passband. The lowpass noise was either Gaussian (reference) or a multiplied noise (comodulated) and had a cut-off frequency of half the bandwidth. The procedure applied in the present study is the same as that described by Hall et al. (1984), where band-pass noise was generated by band-pass filtering a broadband Gaussian or multiplied noise.

In general, the simulated data (right row of Fig. 5.3) are in good agreement with the present experimental data. In particular the total amount of CMR is accounted for by the single-channel model at least for center frequencies of 2 kHz and 4 kHz. However, for $CF = 1$ kHz and for large masker bandwidths the model predicts a CMR effect of only 9 dB, which is 2 dB less than in the averaged data (11 dB). The predicted amount of CMR increases with increasing bandwidth for all center frequencies, which is in agreement with the measured data. However, the absolute value of the predicted threshold deviates in some cases from the observed thresholds. For $CF = 4$ kHz, for example, simulated thresholds are about 4 dB higher than the corresponding measured thresholds, whereas the difference for 1 kHz is less than 3 dB. On the basis of the critical band hypothesis (i.e., assuming peripheral filters with a constant Q-value) one would expect a 6 dB higher threshold for $CF = 4$ kHz than for $CF = 1$ kHz in the reference condition at large masker bandwidths. Indeed, such an effect can be seen in the simulated data, but surprisingly not in the measured data, where the effect amounts to only 4 dB. The deviation might be due to interindividual variances in the growth of peripheral bandwidth as a function of center frequency (e.g. Glasberg and Moore, 1986).

5.4.2 Experiment 2: CMR as a function of modulator and masker bandwidth

Schooneveldt and Moore (1989a) and Carlyon et al. (1989) investigated how the amount of CMR depends on modulator bandwidth. Carlyon et al. (1989) found that the amount of CMR is independent of the modulator bandwidth in the range between 16 and 50 Hz. Schooneveldt and Moore (1989a) measured thresholds for a wider range of modulator bandwidths. They found that for modulator bandwidths larger than 50 Hz, CMR decreases with increasing modulator bandwidth whereas for modulator bandwidths $\leq$ 50 Hz there was no dependence of CMR (for large masker bandwidths) on the modulator bandwidth which is in accordance with the findings of Carlyon et al..

The experiment performed by Schooneveldt and Moore was replicated here for modulator bandwidths Δf_{mod} of 12.5, 50, and 200 Hz at $CF = 2$ kHz. Signal parameters were the same as described in Methods. They are slightly different to those used in Schooneveldt and Moore (1989a): The signal duration was 300 ms instead of 400 ms in the study of Schooneveldt and Moore, and the signals were gated with 50 ms ramps in contrast to 10 ms ramps in the study of Schooneveldt and Moore. Furthermore, the noise masker onset was 150 ms before the signal onset and had a total duration of 600 ms including 10 ms raised cosine ramps whereas Schooneveldt and Moore used continuous noise. Hall et al. (1996) showed that in broadband conditions there is no effect of masker gating even when the signal is gated on and off synchronously with the masker. Thus, differences between the present experimental data and results from Schooneveldt and Moore due to gating effects are not expected at least for large masker bandwidths.

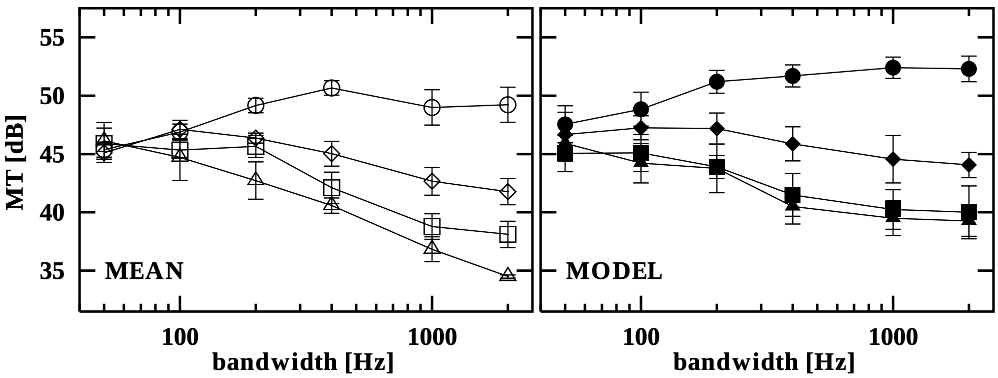

Figure 5.4: *Mean measured data over three subjects (left panel) and simulated thresholds (right panel) for a signal presented in random noise (○) and modulated noise as a function of the masker bandwidth. The modulator bandwidth was 12.5 Hz (△), 50 Hz (□), and 200 Hz (◇), respectively.*

The left panel of Fig. 5.4 shows mean measured data of three subjects for modulator bandwidths Δf_{mod} = 12.5 Hz (△), 50 Hz (□), and 200 Hz (◇), and for the reference condition (○). Thresholds in the comodulated conditions are always lower than in the reference condition. In agreement with the data of Schooneveldt and Moore (1989), the measured CMR-effect decreases with increasing modulator bandwidth. The largest amount of CMR is about 14 dB for Δf_{mod} = 12.5 Hz whereas it is only 7 dB for Δf_{mod} = 200 Hz.

The right panel of Fig. 5.4 shows simulated data. There is again a good agreement between simulated data and measured data. The model shows a reduction in CMR for a modulator bandwidth of 200 Hz which is the same as in the data, whereas the magnitude of the CMR varies little for the modulator bandwidths 12.5 and 50 Hz in agreement with the data. Furthermore, for the smallest masker bandwidths there is no significant threshold difference between the reference condition and the comodulated condition independent of modulator bandwidth. This is also seen in the experimental data. Concerning this point, the present data are not consistent with the data in the literature, where a CMR effect of 4 dB was found for a modulator bandwidth Δf_{mod} = 12.5 Hz and a masker bandwidth of 50 Hz. This difference might be due to a different way of bandlimiting the noise. In the present study the modulator and the noise masker was filtered digitally whereas Schooneveldt and Moore used analog filters. Therefore, the difference may be due to analog filtering and not a real CMR effect (see discussion).

5.4.3 Experiment 3: CMR as a function of signal duration and masker bandwidth

Schooneveldt and Moore (1989a) investigated the effect of signal duration on CMR. Their experiment was replicated here for three signal durations (25, 200, 400 ms), a modulator bandwidth of 12.5 Hz and a center frequency of 2 kHz. Both the masker and signal were gated with 10 ms raised cosine ramps.

Mean measured data over three subjects are shown in the left three panels in Fig. 5.5. Thresholds in the reference condition (○) increase with decreasing signal duration for all masker bandwidths. In the case of wide noise maskers ($\Delta f \geq 1000$ Hz) and a signal duration of 25 ms, for example, thresholds are 9 dB higher than thresholds for a 200 ms signal. This decrease in threshold is in line with experiments on temporal integration (e.g. Florentine et al., 1988), where doubling the signal duration results in a decrease in threshold of about 3 dB. For signal durations of 200 and 400 ms the general shape of the threshold function for the comodulated condition (□) is similar to that found before (experiments 1 and 2), e.g. for large masker bandwidths thresholds are always lower than those in the reference condition whereas no CMR can be found for the smallest bandwidth (50 Hz). The largest CMR is about 3 dB smaller for 200 ms duration than for the 400 ms duration. The finding generally agrees with the data presented by Schooneveldt and Moore (1989a). For a signal duration of 25 ms, however, thresholds in the comodulated conditions are about the same in the two conditions for almost all masker bandwidths. This findings contrasts with the results from Schoonefeldt and Moore.

The right column of Fig. 5.5 shows simulated data. The model accounts qualitatively for the threshold decrease in the reference condition with increasing signal duration.

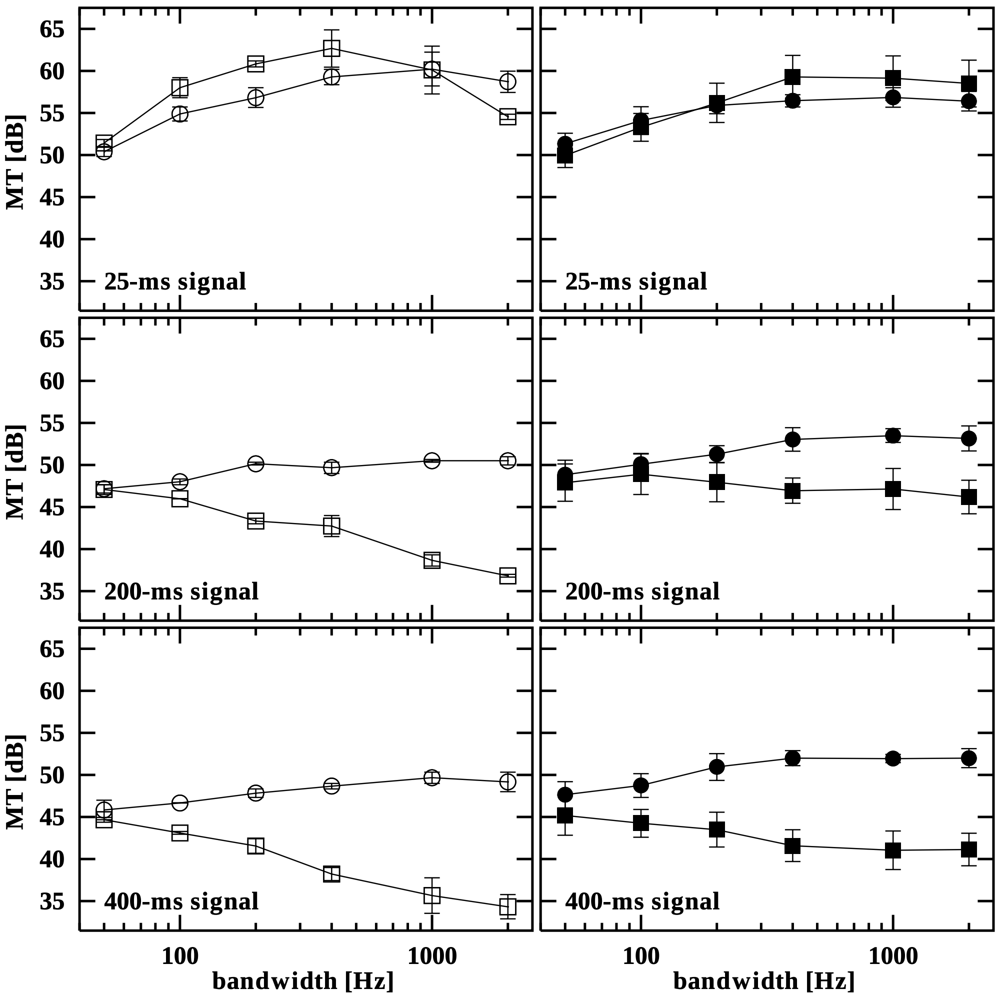

Figure 5.5: *Mean measured thresholds over three subjects (left row of panels) and simulated thresholds (right row of panels) for a signal presented in random noise (○) and modulated noise (modulator bandwidth of 12.5 Hz, □) as a function of the masker bandwidth. Signal duration was either 400 (upper panels), 200 (middle panels), or 25 ms (lower panels). The signal frequency was 2 kHz.*

However the difference between the highest threshold for a duration of 25 ms and the highest threshold for a duration of 400 ms amounts to only 5 dB. This contrasts to the measurements (10 dB). The disability of the model to describe temporal integration was already mentioned in Dau et al. (1996b). The authors explained the shortcoming by the unrealistic behaviour of the feedback loops at the stimulus onset.

The model accounts for the findings that (i) there is CMR for signal durations of 200 and 400 ms and (ii) CMR breaks down at a signal duration of 25 ms. Nevertheless, the predicted maximum CMR for the large durations is smaller (7 dB for CF = 2 kHz and 11 dB for CF = 4 kHz) than the measured CMR (13 dB for CF = 2 kHz and 15 dB for CF = 4 kHz). The smaller amount of CMR in the simulated case in comparison to the simulations in the previous section is mainly a consequence of the shorter ramps of the signal (see discussion).

5.4.4 Experiment 4: Effects of bandlimiting the masker before or after modulation

This section analyzes which factors are responsible for the decrease in CMR with decreasing masker bandwidth Δf. A new experiment was performed where the comodulated masker was generated in a slightly different manner than before (see Verhey and van de Par, 1997). The experimental conditions were equivalent to those in the previous two experiments. The signal frequency was 2 kHz and a modulator bandwidth of 50 Hz was used.

In contrast to the classical generation of comodulated noise, a broadband Gaussian noise was first bandpass filtered and then modulated (e.g., multiplied with a low-pass noise). To obtain the same masker bandwidth Δf as in the previous experiments, the broadband noise was bandlimited to $\Delta f - 2 \cdot \Delta f_{mod}$ prior to modulation, where Δf_{mod} is the modulator bandwidth. Hence, the smallest masker bandwidth that can be achieved is determined by twice the modulator bandwidth $\Delta f = 2 \cdot \Delta f_{mod}$, provided that a sinusoid with random starting phase is used as carrier. The overall level of the masker in the comodulated condition (condition 2) was the same as in the classical comodulated condition (condition 1) with the same masker bandwidth.

Figure 5.6 shows mean measured data of three subjects (TD, KT, JV) [3](left panel) and simulated data (right panel). Circles indicate thresholds in the reference condition, squares indicate thresholds in the "classical" comodulated condition (condition 1) and triangles indicate the thresholds obtained in the new experimental condition (condition 2). For almost all masker bandwidths, thresholds are markedly lower (up to 5 dB) in condition 2 than in condition 1. This difference diminishes for the largest masker bandwidth of 2000 Hz. Simulated thresholds are plotted in the right panel of Fig. 5.6. The model predicts lower thresholds for condition 2 than for condition 1 for masker bandwidths smaller than 1000 Hz, whereas thresholds are the same for both conditions for the largest masker bandwidths. However, some deviations between simulated and measured data occur at small masker bandwidths: In the experimental data, the threshold in the prefiltered comodulated condition (condition 2) tends to slightly increase with decreasing masker bandwidth whereas the model predicts a threshold almost independent of masker bandwidth. However, at small masker bandwidths both intra- and interindividual standard deviations increase with decreasing masker bandwidth in the experimental

[3]In this experiment one of the subjects (KT) did not participate in the previous experiment 1. Therefore small deviations between mean thresholds in Fig. 5.3 and those in this section occur.

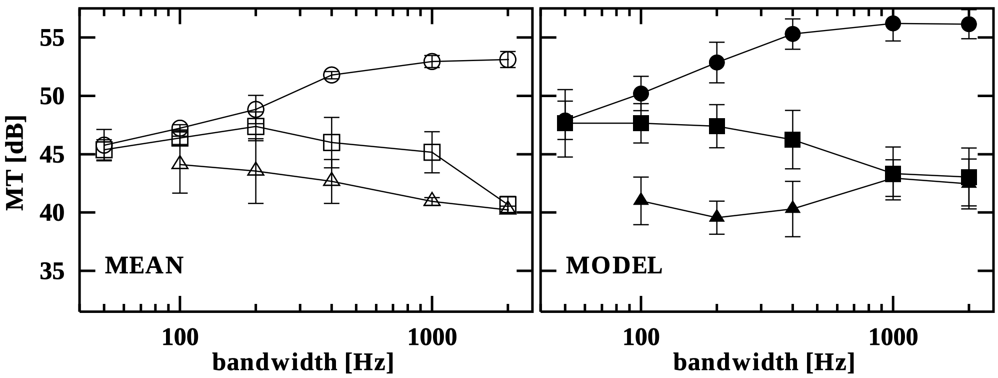

Figure 5.6: *Thresholds for a 4 kHz sinusoidal signal in random noise (○), and in two different modulated noises as a function of masker bandwidth. □ indicate thresholds for the modulated condition, where the masker is first modulated and then bandlimited. This is the classical way to generate modulated noise. △ indicate thresholds, where the masker is first band-pass filtered and then modulated. The left panel show mean measured data over three subjects and the right panel show simulated data.*

data. This is in line with the increasing standard deviation with decreasing masker bandwidth in the simulations for condition 2. Hence, the deviation between experimental and theoretical data is not significant.

5.4.5 Experiment 5: Effects of varying masker modulation depth

Moore (1992) reviewed the factors that influence the occurrence of CMR. He proposed that CMR is large for high modulation depths in the comodulated condition. This assumption is based on findings from Moore and Jorasz (1991) and Fantini and Moore (1992). In their studies the task was to detect a change in modulation depth between two spectrally separated carriers. Hence, these experiments belong to the second class of CMR-experiments (see introduction). Concerning the first class of experiments, however, there is no comparable study on the effect of masker modulation depth on CMR.

In the present study the masker modulation depth is varied in the comodulated condition by adding a dc-component to the low-pass noise modulator. The procedure to generate the comodulated-noise masker was the same as described in Methods. Figure 5.7 shows masked thresholds for a 2-kHz sinusoidal signal in the presence of a 1000-Hz wide bandpass noise-masker as a function of the amount of added dc-component. The magnitude of the dc-component is expressed as the difference in level (dB) between the dc-component and the root-mean-square value of the low-pass noise-modulator [4]. The

[4]The difference in level between the dc-component and the root-mean square of the low-pass noise-modulator was $20 * \log(dc/rms_{lp})$, where rms_{lp}, the root mean square value of the low-pass noise, was 0.316 and the dc component was 0.1, 0.2, 0.5, 0.8, 1.0, or 2.0.

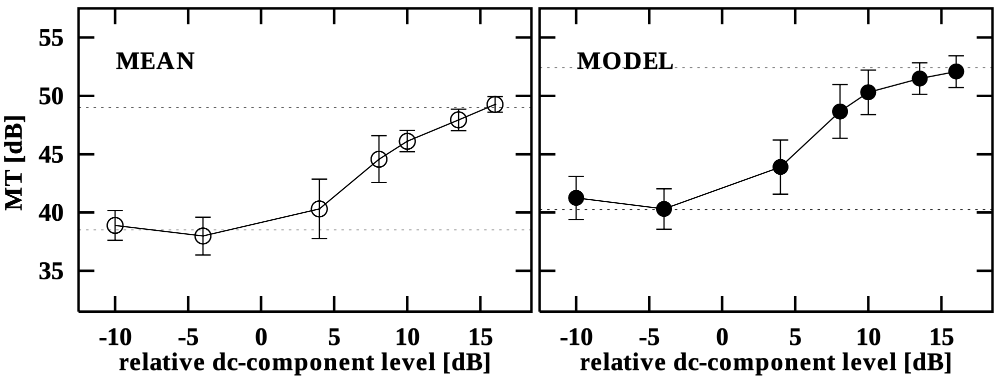

Figure 5.7: *Mean measured data (right panel) and simulated data (left panel) for a sinusoidal signal in the presence of a 1000 Hz modulated band-pass noise-masker as a function of the added dc-component to the modulator. In addition thresholds for random noise (upper line) and modulated noise (lower line), where no dc-value is added (experimental condition as in Fig. 1) are shown.*

left panel of Fig. 5.7 shows measured data averaged across four subjects (TD, KT, KS, JV). The right panel shows simulated data. The dotted lines show threshold for the reference condition (upper line) and the "classical" comodulated condition (lower line), where no dc-component was added for a masker bandwidth of 1000 Hz. For dc values of -10 dB and -3 dB thresholds first remain at a constant level which corresponds to the threshold in the classical comodulated condition. They increase for dc-values larger than -3 dB. The thresholds converge at the threshold in the reference condition for large dc-values, leading to a sigmoidal shape of the threshold curve. Again the simulated data are in good agreement with experimental results.

5.4.6 Experiment 6: CMR for square-wave modulation

In a recent study on CMR, Hall et al. (1996) presented an experiment belonging to the first class of CMR experiments (see introduction). In contrast to most experiments of this class of CMR experiments they used square-wave modulation instead of lowpass modulation. In that case, the CMR effect was significantly larger than for lowpass-modulators. For example, Hall et al.'s data for 10-Hz square-wave modulation and a signal duration of 400 ms showed a CMR effect of 40 dB for large masker bandwidths compared to about 15 dB for a lowpass-noise modulator with a modulator bandwidth $\Delta f_{mod} = 12.5$ Hz in Schooneveldt and Moore's experimental results (1989a). A smaller difference of about 5 to 10 dB is also apparent for a modulation frequency of 40 Hz, where the total CMR amounts to 18 dB.

Of course, a direct comparison between Schooneveldt and Moore's experimental results and Hall et al.'s data is difficult, because they did not use the same experimental setup. We performed an experiment to directly relate both types of experiments to each other. The signal configuration was the same as in Experiment 1 except for the modulating waveform, which consists of a square-wave with a repetition rate of 50 Hz and a duty cycle of 50:50. The modulator was dc-shifted, i.e., it only assumed the values of 0 and 1. The carrier signal was again a sinusoid at 2 kHz with 300 ms duration.

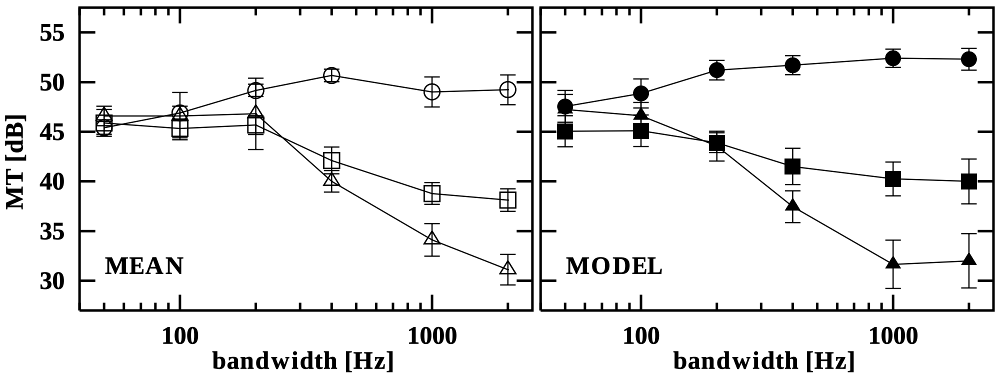

Figure 5.8: *Mean data over three subjects (right panel) and simulated data (left panel) showing thresholds for a signal presented in random noise (○), in modulated noise with a 50 Hz lowpass-noise modulator (□), and in modulated noise with a 50 Hz square-wave modulator (△) as a function masker bandwidth.*

Figure 5.8 shows masked thresholds in the modulated condition with the square-wave modulated masker (△). The left panel shows mean measured data of four subjects. The right panel shows simulated data. In addition, for comparison, the data for the lowpass-noise modulator with cut-off frequency equal to the square-wave frequency (□) and the thresholds for the unmodulated condition (○) are replotted from Fig. 5.2. The measured threshold curve for square-wave modulation is equal to or lower than for the lowpass-noise modulators. The difference between the two conditions increases with increasing bandwidth up to about 8 dB. Square-wave modulators produce a larger CMR (up to about 20 dB) than lowpass-noise modulators. This is in line with findings by Hall et al. (1996)(see above).

Apart from the fact that the simulated data are slightly shifted towards higher thresholds, they are in good agreement with the measured data. The model accounts for the increasing difference between the two conditions with increasing masker bandwidth. For large bandwidths ($\geq$ 1000 Hz) the simulated CMR for the square-wave modulator is 18 dB. Thus, the total amount of CMR is accounted for by the model.

5.5 DISCUSSION

The role of within-channel cues in this class of CMR experiments (characterized by a single bandpass noise centered at the signal frequency) has often been discussed. To quantify the role of within-channel cues, in the present study simulations with a single-channel modulation-filterbank model were performed and compared to experimental results. The most important finding is that in most cases the total amount of masking release can be predicted by a single-channel analysis, provided that the temporal envelope within this channel is further considered in an appropriate way.

5.5.1 Role of modulation processing

In the comodulated condition, the envelope of the masker shows distinct minima because the noise is overmodulated, i.e. the envelope spends a large proportion of time at values near zero. Adding a signal to the modulated noise will smooth out the sharp minima, as already mentioned by Schooneveldt and Moore (1989a). This yields a decrease in modulation depth, which is detected by the model presumably in a similar way as in human listeners. Since the minima are very distinct and mark only very short time intervals, high modulation rates must be processed by the model, i.e., a good temporal resolution is required (as already stated by Buus, 1985). The role of high modulation rates is shown in Fig. 5.9. The simulated CMR effect is shown for a masker bandwidth of 1000 Hz, a modulator bandwidth of 50 Hz, and a signal frequency of 2 kHz as a function of the highest modulation center frequency of the modulation filterbank used in the model. All previously presented simulations were performed with an upper cut-off frequency of a quarter of the signal frequency CF (center frequency of the peripheral filter), i.e., in this case 500 Hz for $CF = 2$ kHz. The predicted CMR is independent of the highest modulation center frequency (CF_{mod}) for $CF_{mod} \geq 250$ Hz. For $CF_{mod} < 250$ Hz, the predicted CMR decreases with decreasing CF_{mod}. The CMR effect almost completely disappears when only a modulation lowpass filter as implemented in Dau et al. (1996a) is used (first data point in Fig. 5.9).

As mentioned in the literature (Moore, 1992; Moore and Jorasz, 1991, 1996), CMR occurs when the added signal produces an increase in level in the valleys of the masker. This is possible if the modulation depth of the masker is high. In experiment 5 the effect of modulation depth was tested explicitly (see Fig. 5.7). It was shown that if the modulation depth is decreased, the CMR is decreased and diminishes for large added dc-components. This is consistent with the hypothesis that changes in modulation depth are the reason for the masking release in this kind of experiment. When a dc-component is added, the signal level has to be higher to produce a decrease in modulation depth. When the dc-component is very large, the change in overall energy produced by adding the signal will be the strongest cue, whereas changes in modulation depth will play a secondary role. This latter strategy is equivalent to the strategy used in the reference condition. As a consequence, no CMR will be observed, as shown in the experimental results.

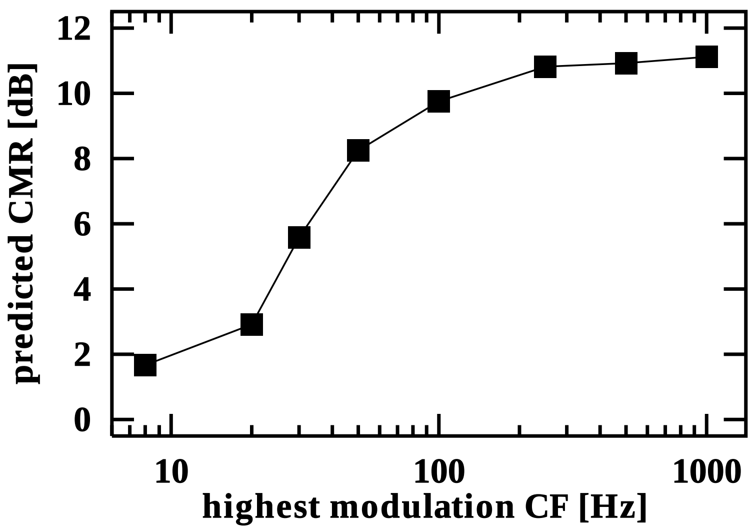

Figure 5.9: *Simulated CMR for a masker bandwidth of 1000 Hz, a modulator bandwidth of 50 Hz, and a signal frequency of 2 kHz. Parameter is the highest modulation center frequency of the modulation filters.*

If the comodulated masker consists of a square-wave modulated noise (see experiment 6 and Fig. 5.8), envelope amplitudes spend more time near zero in comparison with the classical comodulated condition that employ a lowpass-modulator. As a consequence, and in agreement with the model predictions, the masking release is larger than in the classical CMR experiments. This point of view agrees with Zwicker and Schorn's (1982) assumption that the reduced masking in square-wave modulated noise compared to continuous noise (reference) can be interpreted in terms of temporal resolution. An across-channel process, as stated in the literature (Buus, 1985; Hall et al., 1996) seems not to be necessary in the class of experiments considered here.

5.5.2 Role of the critical-band filter

It was proposed in the literature (Carlyon et al., 1989) that the across-frequency proportion of the masking release could be estimated by subtracting out the masking release for a masker bandwidth equal to the auditory filter bandwidth at the signal frequency. However, as mentioned in the introduction, this procedure is somewhat problematic because it postulates a clear distinction between within-channel and across-channel effects at the value of the critical bandwidth (Hall et al., 1996). The present study shows that this approach is not able to predict "across-channel" CMR properly: Even a single-channel model operating on only one critical band predicts a decrease in thresholds for masker bandwidths larger than the critical bandwidth in the comodulated condition (see, e.g., Fig. 5.3). The predicted decrease in threshold beyond the critical bandwidth is consequence of the auditory filters incorporated in the model. Although frequency components of the masker outside the critical bandwidth are strongly attenuated, they still contribute to the output of the auditory filter. To quantify this effect, simulations were performed with a model that employs a rectangular filter with the same bandwidth instead of the gammatone filter. Figure 5.10 shows simulated data on the basis of the modified model

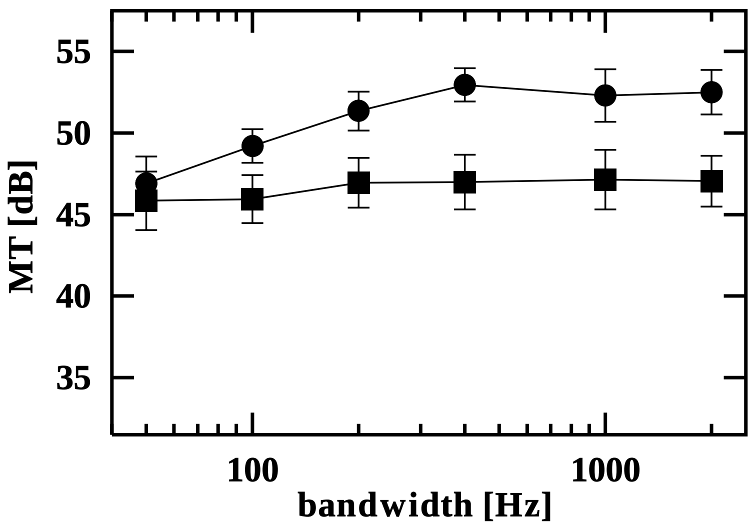

Figure 5.10: *Simulated data for a 2 kHz sinusoidal signal in random noise (circles) and comodulated noise (squares). In this simulation instead of a gammatone filter a rectangular filter with the same bandwidth was applied.*

for $CF = 2\,\mathrm{kHz}$. Circles represent thresholds in the reference condition whereas squares represent thresholds in the comodulated condition. In the reference condition thresholds increase with increasing masker bandwidth up to the critical bandwidth and then stay constant whereas in the comodulated condition thresholds are independent of masker bandwidth. As expected, there is no increase in CMR for masker bandwidths larger than the critical bandwidth. The predicted CMR of 6 dB for large masker bandwidths is 5 dB smaller than the simulated CMR effect with the original model (c.f. Fig. 5.2).

5.5.3 Role of filtering the masker

In the classical comodulated condition the masker is bandlimited after modulation. As argued in the literature on modulation detection (Eddins, 1993; Strickland and Viemeister, 1997), spectral filtering after modulation decreases the modulation depth of the stimuli. Schooneveldt and Moore (1989a) discussed this effect with respect to cases when the modulator bandwidth is equal or greater than half the masker bandwidth. They argued that frequency components of the modulator greater than half the bandwidth are filtered out and no longer contribute to the modulation. They concluded that conditions where the modulator bandwidth was equal or greater than half the masker bandwidth are equivalent.

However, also for modulator bandwidths smaller than half the masker bandwidth the modulation depth still will be reduced by filtering after modulation. The hypothesis in the present study is that the reduction in modulation depth due to filtering after modulation is the reason for the increase in threshold with decreasing masker bandwidth in the comodulated condition. It is based on the assumption that the reduced thresholds in the comodulated condition compared to those in the reference condition are a consequence of changes in modulation depth when the signal is added. To test our hypothesis, experiment 4 was performed (see Fig. 5.6) where the masker was first bandlimited and then modulated. It was shown that the slope of the threshold curve is markedly reduced when the masker is modulated after filtering. This provides further evidence that (i) thresholds in the comodulated condition are strongly affected by the masker modulation depth, (ii) that the decrease in threshold with increasing masker bandwidth in the classical comodulated condition is mainly due to the method of generating the comodulated masker.

5.5.4 Relation to modulation detection experiments

For large masker bandwidths ($\leq$ 800 Hz) CMR decreases for modulator bandwidths Δf_{mod} larger than 50 Hz whereas CMR is more or less independent of modulator bandwidth for $\Delta f_{mod} \geq 50$ Hz (experiment 2 ; Schooneveldt and Moore, 1989a). Schooneveldt and Moore compared this result with the data on amplitude modulation detection in noise (e.g., Viemeister, 1979). They argued that CMR is largest for modulator bandwidths smaller than or equal 50 Hz because the auditory system is most sensitive to amplitude modulation in this modulation frequency region. However, a direct comparison of the data on modulation detection and results from CMR experiments is difficult for several reasons: First, Viemeister measured modulation detection for sinusoidal modulation, whereas in CMR experiments generally noise modulators were used, i.e., the imposed modulator is deterministic in the modulation detection task whereas it is random in the CMR experiment. Second, whereas in modulation detection experiments an imposed modulation is to be detected, in CMR experiments the sinusoidal signal effectively changes the modulation depth of the modulator. However, it is difficult to quantify the modulation depth of a carrier which is overmodulated with a noise modulator. A more appropriate description of the modulated noise may be given by the envelope distribution, a technique which was first proposed by Schooneveldt and Moore (1989a). They found that for a modulator bandwidth of 200 Hz the envelope spends a smaller proportion of time at values near 0 compared to a modulator bandwidth of 50 Hz. That means less minima are partially filled in by the signal. Thus, the reduction in overall modulation depth will be smaller than in the 50 Hz modulator condition. As a consequence, our model predicts a smaller CMR effect in agreement with the data.

5.5.5 Role of the carrier duration

The CMR effect depends on the duration of the signal (experiment 3; Schooneveldt and Moore, 1989a). With decreasing signal duration the CMR effect decreases. The current model predicts this effect because the time for which the modulation depth is reduced due to the presence of the signal is reduced. Schooneveldt and Moore (1989a) argued that the auditory system needs a reasonable length of time to sample the signal, if within-channel cues are exploited. They concluded that for durations greater than 100 ms within-channel processes might contribute to the masking release in CMR experiments. In contrast, for durations $\leq$ 100 ms they assumed that CMR is a consequence of a "pure" across-channel process where within-channel cues do not play any role. The present study cannot rule out this hypothesis. However, not such a CMR of about 5 dB for a signal duration of 25 ms as measured by Schooneveldt and Moore was found in the present experiments. In agreement with the model predictions, the present data show no significant CMR for a signal duration of 25 ms. Nevertheless, simulated CMR for signal duration of 200 ms and 400 ms is smaller than the measured CMR. The deviations of the simulated results from the experimentally derived results can be explained in the same way as was already done with respect to simulations to temporal integration by the model (Dau et al., 1996b). Because the adaptation loops are very sensitive to fast changes in the input, the short signal ramps in the experiment (10 ms) will cause very strong oscillations at the output of this stage when the signal is switched on. As a consequence, the template for the simulated experimental runs are dominated by an overshoot at the first few milliseconds of the signal. This unrealistic weighting of signal information over time can be reduced if longer signal ramps are used. In fact, as can be seen in section 5.5.2, the whole CMR effect can be predicted by the model, if longer ramps (e.g., 50 ms) are used.

5.5.6 Modulation-filterbank concept

The modulation filterbank model which was originally developed to describe modulation detection and modulation masking experiments was used in the present study. It was assumed that within-channel cues in CMR are based on the ability of the auditory system to detect changes in modulation depth. The modulation filterbank model (Dau et al. 1997a+b) differs considerable in its structure from the model of modulation detection by Viemeister (1979). The model of Viemeister consists of a broad prediction filter (with a bandwidth Δf = 2000 Hz) which is followed by a half-wave rectification and a lowpass filter with a cut-off frequency of 64 Hz. As decision variable Viemeister suggested the ac-coupled root-mean square of the output of the low-pass filter. Based on the Viemeister model, Berg (1996) presented simulated data with a slightly modified model for the second class of CMR experiments. This model derives the calculations of the decision statistic from the amplitude spectrum of the output of the leaky integrator. Berg showed that (i) CMR can be predicted by a model based on the envelope amplitude spectrum, (ii) the disruption of the CMR effect by adding a single tone to the stimuli can be simulated by the model. Unfortunately, the distance of comodulation bands in his CMR experiment is in the range of the critical bandwidth of the peripheral filter at the signal frequency. Thus, it is generally assumed that within-channel processes might contribute to the masking release in the CMR experiment performed by Berg.

The present study did not use the Viemeister model to describe modulation detection because of several reasons: (i) First, the Viemeister model is not adequate to simulate within-channel cues. The term "within-channel cue" assumes that the auditory system extracts envelope information from the temporal output of each peripheral filter, whereas in the Viemeister model the concept of peripheral filtering is not incorporated. (ii) Second, as a consequence, the Viemeister model would fail to predict the present data (first class of CMR experiments), because the thresholds in the reference condition are assumed to be a function of the energy in the peripheral filter at the signal frequency. In contrast to the Viemeister model, the modulation filterbank model is able to predict threshold curves both for the reference and comodulated condition because it assumes that the modulation analysis is performed in parallel on the output of each peripheral channel.

5.6 CONCLUSION

A single-channel model accounts quantitatively for several aspects of CMR. Specifically, the effects of varying masker bandwidth, signal duration, modulator bandwidth, modulator type and center frequency are predicted reasonably well with an appropriate single-channel analysis. Modulation filters with a center frequency up to at least 250 Hz are necessary to simulate the whole amount of CMR. In agreement with assumptions in the literature, high modulation depths of the masker are necessary to obtain large CMR effects. The decrease in threshold for masker bandwidths larger than the critical bandwidth can be predicted because the model does not use unrealistic rectangular filters. Hence, frequency components outside the critical band, although strongly attenuated by the peripheral filter, still affect the performance of the model. Filtering the masker after modulation reduces the modulation depth of the masker especially for masker bandwidths which are comparable to the modulator bandwidth. The decrease in threshold with increasing masker bandwidth in the comodulated condition is therefore mainly a consequence of the technique to generate the masker. In some conditions, the predicted CMR is slightly smaller than in the experimental results. This difference between measured and simulated results might be a consequence of a “true” across-channel process. However, this effect amounts to a maximum size of only about 3 dB.

Chapter 6

General Conclusions

Whereas powerful models concerning the spectral processing of the auditory system exist, the processing of temporally varying sounds is often unclear. Recently, a model of modulation processing was presented by Dau et al. (1997a+b) to account for modulation detection and masking data. In Chapter 4 of this thesis, it was investigated, if the model can also predict spectral masking data. It was shown that the model accounts apart from frequency selectivity and off-frequency listening effects in notched-noise experiments also for the effect of "asymmetry of masking", where it is observed that thresholds for test-signals with bandwidths larger than the masker bandwidth are much lower than those for the reversed condition, although all signal bandwidths are smaller than a critical bandwidth. A power-spectrum model, where only the energy in each critical band is analyzed, would predict the same threshold for both conditions. In contrast to the power-spectrum model, the model presented by Dau et al. (1997a+b) is able to use the inherent higher modulation frequencies of the test signal as an additional cue, when the test-signal bandwidth is larger than the masker bandwidth. Another phenomenon, which can not be explained by a power spectrum model are the differing masking properties of modulated and unmodulated masker, which are observed in CMR-experiments. In the literature it is generally assumed that the release is a consequence of an across-channel process, where the subjects make use of the information that the envelope fluctuations of the masker in peripheral filters at different center frequencies is coherent. However, some authors argued that some of the release might be due to changes in the envelope statistics within the critical band centered at the signal frequency (within-channel cues). To investigate the role of within-channel cues in CMR-experiments, in Chapter 5, simulations with a single-channel version of the above processing model were performed, where only the information out of the peripheral filter centered at the signal frequency were processed. It was shown that the masking release in some of the "classical" CMR-experiments can be predicted, when an appropriate analysis is performed in one peripheral channel. The simulations in Chapter 5 clarify the role of within-channel cues quantitatively. In the experimental conditions only 3 dB of the total measured masking release of 15 dB could be contributed to real across-channel processes. These results are basis for further modeling efforts towards a model incorporating across-channel processes.

Whereas in Chapters 4 and 5, temporal effects in spectral *masking* were investigated, temporal aspects of spectral integration of *loudness* were measured in Chapters 2 and 3. Since methodological factors often influence loudness measurements, it is important to quantify and minimize these influences. It was shown, that the influence of methodological factors can not be separated from the effect of the variation of the physical signal parameters when a simple adaptive procedure (with non-interleaved tracks) is used. When an adaptive procedure with interleaved tracks is used, the effect of methodological factors is minimized. Loudness was measured as a function of signal bandwidth and duration with the optimized procedure. It was shown that loudness summation depends on signal duration. It is markedly increased for 10-msec signal, whereas it is almost the same for 100 and 1000-msec signals. Present loudness models of time varying sounds can not account for the experimental data, since their spectral-analysis stage acts nearly instantaneously and duration-independent. A model which contains a duration-dependent compression should in principle account for the present data. The detection model presented by Dau et al. (1996a+b, 1997a+b) assumes a nonlinear compression behavior of the auditory system, that incorporates certain time constants ("adaptive compression"). However, a loudness model including this compression stage would predict a smaller effect for short signals than for longer signals. This contrasts with findings in Chapter 3. Thus, the model presented by Dau et al. can not be easily extended to a loudness model for time varying signals. The findings in Chapter 3 indicate that a new modeling approach is needed to predict the loudness of temporally varying sounds. Further experiments should be performed to investigate how the the loudness summation for different signal durations changes with level. In addition, experiments on temporal aspects of loudness summation with hearing-impaired subjects could help to understand the underlying mechanism.

Appendix A

Equal-loudness level surface

The upper panel of Fig. A.1 illustrates how the combination of experimental results on temporal integration of loudness for different bandwidths with results on loudness summation for stationary signals indirectly give information about loudness summation for short signals (line A–D in Fig. A.1).

First, for a specified level of the short signal the level difference between equally loud short (5 ms) and long (200 ms) white-noise signals is derived from Fig. 3 in Florentine et al. (1996) (line A–B in Fig. A.1). At a level of $\text{Ł}_{short}(noise) = 50\,\text{dB}$ for the short white-noise signal, the level difference between an equally loud short and long noise signals is 12 dB, i.e., the level of the equally loud long noise signal is 38 dB. At this level, for stationary signals Zwicker et al. (1957) measured a level difference of about 12 dB between a broadband signal and an equally loud signal with a subcritical bandwidth (line B–C). Therefore, to produce equal loudness with the long noise signal, the level of a long *tone* ($\text{Ł}_{long}(tone)$) has to be 50 dB. Because the level differences are shown as a function of the level of short signal, the level of an equally loud short tone (line C–D) can not be derived directly from Fig. 3 in Florentine et al. (1996). Instead, it is necessary to find a level of the short tone $\text{Ł}_{short}(tone)$, where the level difference yields 50 dB for the level of the long tone ($\text{Ł}_{long}(tone)$). This is the case for $\text{Ł}_{short}(tone) = 68\,\text{dB}$. Now, an equal-loudness level square is constructed (lower panel of Fig. A.1) which can be interpolated to an equal-loudness level surface. On this surface, it is possible to calculate the level difference between a short white-noise signal of 50 dB and an equally loud short tone ($\text{Ł}_{short}(tone) - \text{Ł}_{short}(noise)$). It amounts to 18 dB (line A–D).

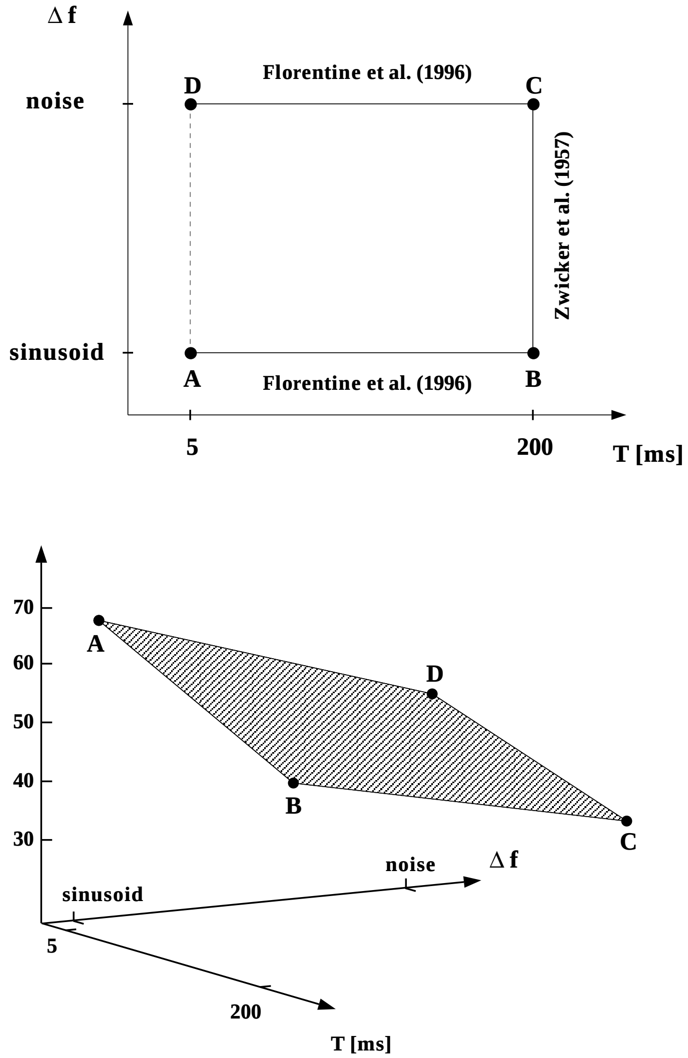

Figure A.1: *Upper panel: Illustration for deriving loudness summation for short signals from experimental results on temporal integration of loudness and loudness summation for stationary signals. Lower panel: Equal loudness level surface for a 5-msec white-noise signal with a level of 50-dB SPL.*

Appendix B

Spectral masking at different center frequencies

In addition to experiment A described in Chapter 4, measurements and simulations of spectral masking with a symmetric notched noise masker at different center frequencies were performed. Apparatus, procedure, and subjects were the same as in experiment B described in Chapter 4. The test signal was a 400-msec long tone. The signal frequency was 1 kHz, 2 kHz, or 4 kHz. It was gated with 200-msec raised-cosine ramps. The test-signal was temporally centered in the masker, which consisted of a pair of 500-msec long 800-Hz wide frozen-noise maskers, one below and one above the signal frequency. To generate the masker bands, a broadband noise was digitally filtered by setting the magnitude of the Fourier coefficients to zero outside the desired passbands (FFT length = 16000 samples, sampling frequency = 32 kHz). Masked thresholds were measured as a function of the relative distance $\Delta f/f$ from the upper edge of the lower bandpass noise to the signal frequency f. The relative distance $\Delta f/f$ was 0, 0.05, 0.1, 0.2, 0.3, and 0.4, respectively. For a center frequency of 1 kHz, only relative distances in the range from 0 to 0.2 were used. The simulations were performed with a multi-channel version of the model. In experiment A it was shown, that high modulation filters do not contain important information that contributes to signal detection in this kind of experiment. Therefore, only the lowpass-filtered envelope of each peripheral filter was processed. Figure B.1 shows measured (open symbols) and simulated data (filled symbols). Clear individual differences are apparent in the measured data. There are differences in threshold of 2 to 8 dB across subjects. This is in agreement with the results from the literature (e.g., Moore et.al., 1995). For the subject td the slope of the threshold curve depends on center frequency. The slope increases with increasing frequency. This does not hold for the other two subjects. The slope of the threshold curves becomes shallower for $\Delta f/f$ larger than 0.2 for the subject jt at 2 kHz, and for the subject jv at 4 kHz, respectively. This effect is probably caused by different absolute thresholds at different frequencies. The model accounts for the general shape and dynamic range of the threshold functions. The slope of the simulated curves is independent from signal frequency. In the model, the absolute threshold was assumed to be the same at the three signal frequencies.

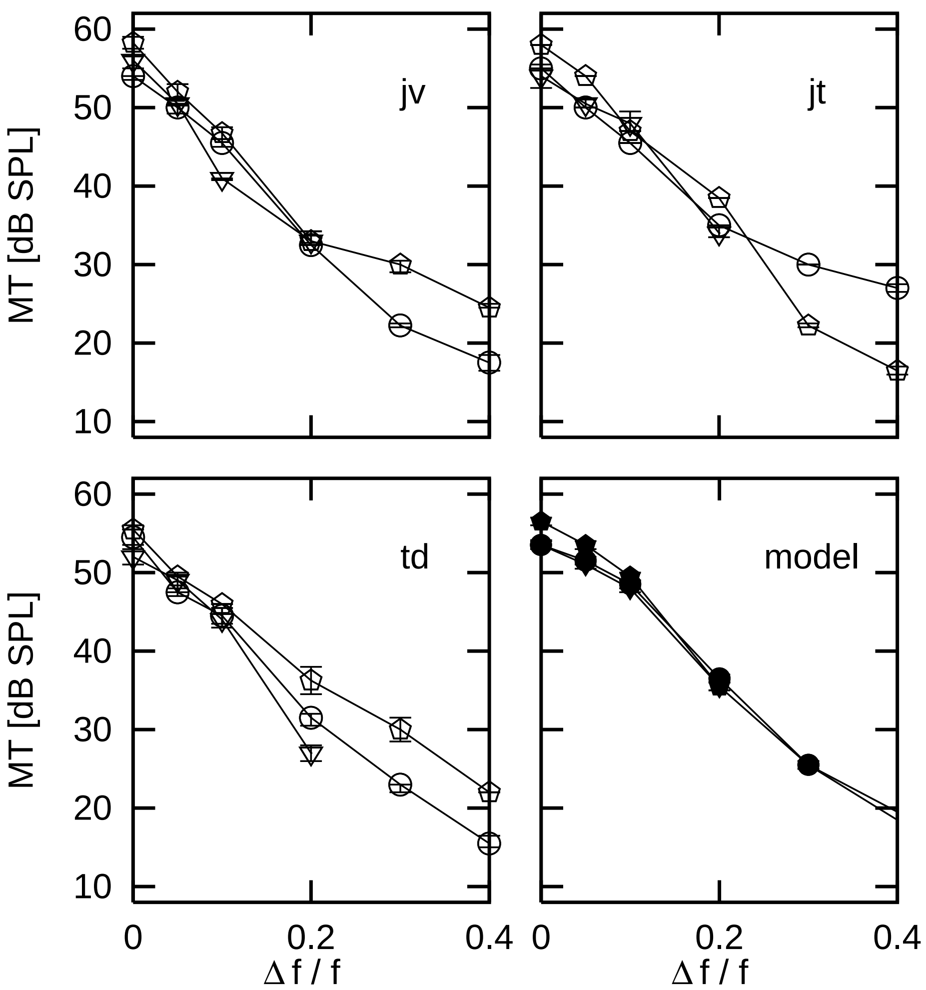

Figure B.1: *Masked thresholds for a test tone presented in a symmetric notched noise masker as a function of the relative distance* $\Delta f/f$ *from the signal frequency* f *to the upper edge of the lower noise band. The panels with open symbols indicate measured thresholds for three different subjects. Each panel shows data for three frequencies: 1 kHz (triangles), 2 kHz (circles) and 4 kHz (pentagons). The lower right panel (filled symbols) shows simulated thresholds with the present model.*

References

Bacon, S.P., and Viemeister, N.F. (**1985**):
"The temporal course of simultaneous tone-on-tone masking,"
J. Acoust. Soc. Am. 78, 1231–1235.

Bacon, S.P., and Moore, B.C.J. (**1986a**):
"Temporal effects in simultaneous pure-tone masking: Effects of signal frequency, masker/signal frequency ratio, and masker level,"
Hear. Res. 23, 257–266.

Bacon, S.P., and Moore, B.C.J. (**1986b**):
"Temporal effects in masking and their influence on psychophysical tuning curves,"
J. Acoust. Soc. Am. 80, 1638–1645.

v. Békésy, G. (**1942**):
"Über Schwingungen der Schneckentrennwand beim Präparat und Ohrmodell,"
Akustische Zeits. 7, 173–186.

v. Békésy, G. (**1943**):
"Über die Resonanzkurve und Abklingzeit der verschiedenen Stellen der Schneckentrennwand,"
Akustische Zeits. 8 , 66–76.

v. Békésy, G. (**1947**):
"The variation f phase along the basilar membrane with sinusoidal vibrations,"
J. Acoust. Soc. Am. 19, 295–300.

v. Békésy, G. (**1949a**):
"Elasticity of cochlear partition,"
J. Acoust. Soc. Am. 21, 227–232.

v. Békésy, G. (**1949b**):
"The vibration of the cochlear partition in anatomical preparations and in models of the inner ear,"
J. Acoust. Soc. Am. 21, 233–254.

Berg, B.G. (**1996**):
"On the relation between comodulation masking release and temporal modulation transfer functions,"
J. Acoust. Soc. Am. 100, 1013–1023.

Boone, M.M. (**1973**):
"Loudness measurements on pure tone and broad band impulsive sounds,"
Acustica 29, 198–204.

Bos, C.E., and de Boer, E. (**1966**):
"Masking and Discrimination,"
J. Acoust. Soc. Am. 39, 708–715.

Buus, S. (**1985**):
"Release from masking caused by envelope fluctuations,"
J. Acoust. Soc. Am. 78, 1958-1965.

Buus, S., Florentine, M., and Poulsen, T. (**1997**):
"Temporal integration of loudness, loudness discrimination, and the form of the loudness function,"
J. Acoust. Soc. Am. 101, 669–680.

Buus, S., Florentine, M., and Müsch, H. (**1998**):
"Loudness function for tones at low levels derived from loudness summation,"
in *Psychophysical and Physiological Advances in Hearing*, XIth International Symposium on Hearing, ed. Palmer, A. et al. 449–457.

Cacace, a.T., and Margolis, R.H. (**1985**):
"On the loudness of complex stimuli and its relationship to cochlear excitation,"
J. Acoust. Soc. Am. 78, 1568–1573.

Carlyon, R.P., Buus, S. and Florentine, M. (**1989**):
"Comodulation Masking Release for three types of modulator as a function of modulation rate,"
Hear. Res. 42, 37–46.

Cohen, M.F. (**1991**):
"Comodulation masking release over a three octave range,"
J. Acoust. Soc. Am. 90, 1381–1384.

Dau, T., Püschel, D. and Kohlrausch, A. (**1996a**):
"A quantitative model of the "effective" signal processing in the auditory system: I. Model structure,"
J. Acoust. Soc. Am. 99, 3615–3622.

Dau, T., Püschel, D. and Kohlrausch, A. (**1996b**):
"A quantitative model of the "effective" signal processing in the auditory system:

II. Simulations and measurements,"
J. Acoust. Soc. Am. 99, 3623–3631.

Dau, T., Kollmeier, B, and Kohlrausch, A. (**1997a**):
"Modeling auditory processing of amplitude modulation. I. Modulation detection and masking with narrowband carriers,"
J. Acoust. Soc. Am. 102, 2892–2905.

Dau, T., Kollmeier, B, and Kohlrausch, A. (**1997b**):
"Modeling auditory processing of amplitude modulation. I'I. Spectral and temporal integration in modulation detection,"
J. Acoust. Soc. Am. 102, 2906–2919.

Derleth, R.P., Dau, T., and Kollmeier, B. ():
"Modeling frequency selectivity with a perception model,"
in preparation.

de Boer, E. (**1962**):
"Note on the critical bandwidth,"
J. Acoust. Soc. Am. 34, 985–986.

de Boer, E. (**1966**):
"Intensity discrimination of fluctuating signals,"
J. Acoust. Soc. Am. 40, 552–560.

Eddins, D. (**1993**):
"Amplitude modulation detection of narrow-band noise: Effects of absolute bandwidth and frequency region,"
J. Acoust. Soc. Am. 93, 470–479.

Eddins, D.A. and Wright, B.A. (**1994**):
"Comodulation masking release for single and multiple rates of envelope fluctuation,"
J. Acoust. Soc. Am. 96, 3432–3442.

Fantini, D.A. and Moore, B.C.J. (**1992**):
"Comodulation Masking Release (CMR) and Profile Analysis: the Effect of Varying Modulation Depth,"
in Advances in Biosciences, Vol. 83, pp. 479–485.

Fantini, D.A., Moore, B.C.J. and Schooneveldt, G.P. (**1993**):
"Comodulation masking release (CMR) as a function of type of signal, gated or continuous masking, monaural or dichotic presentation of flanking bands, and center frequency,"
J. Acoust. Soc. Am. 93, 2106–2115.

Fletcher, H. (**1940**):
"Auditory patterns,"
Rev. Mod. Phys. 12, 47–61.

Fletcher, H., and Munson, W.A. (**1933**):
"Loudness, its definition, measurement and calculation,"
J. Acoust. Soc. Am. 5, 82–108.

Florentine, M., Fastl. H. and Buus, S. (**1988**):
"Temporal integration in normal hearing, cochlear impairment, and impairment simulated by masking,"
J. Acoust. Soc. Am. 84, 195–203.

Florentine, M., Buus, S., and Poulsen, T. (**1996**):
"Temporal integration of loudness as a function of level,"
J. Acoust. Soc. Am. 99, 1633–1644.

Glasberg, B. R. and Moore, B.C. J. (**1986**):
"Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments,"
J. Acoust. Soc. Am. 79, 1020–1033.

Gustafsson, B. (**1974**):
"The loudness of transient sounds as a function of some physical parameters,"
J. Sound Vib. 37, 389–398.

Gabriel, B. (**1996**):
Equal-Loudness Level Contours: Procedures, Facts and Models
Shaker, Aachen.

Gabriel, B., Kollmeier, B. and Mellert, V. (**1997**):
"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, 670–683.

Greenwood, D.D. (**1961**):
"Auditory masking and the critical band,"
J. Acoust. Soc. Am. 33, 484–502.

Greenwood, D.D. (**1971**):
"Aural combination tones and auditory masking,"
J. Acoust. Soc. Am. 50, 502–543.

Hatch, D.R., Arne, B.C. and Hall, J.W. (**1995**):
"Comodulation Masking Release (CMR): Effects of gating as a function of number of flanking bands and masker bandwidth,"
J. Acoust. Soc. Am. 97, 3768–3774.

Haggard, M.P., Hall, J.W. and Grose, J.H. (**1990**):
"Comodulation masking release as a function of bandwidth and test frequency,"
J. Acoust. Soc. Am. 88, 113–118.

Hall, J.L. (**1997**):
"Asymmetry of masking revisited: Generalization of masker and probe bandwidth,"
J. Acoust. Soc. Am. 101, 1023–1033.

Hall, J.W., Haggard, M.P. and Fernandes, M.A. (**1984**):
"Detection in noise by spectro-temporal pattern analysis,"
J. Acoust. Soc. Am. 76, 50–56.

Hall, J.W., Grose, J.H. and Haggard,M.P. (**1989**):
"Effects of flanking band proximity, number, and modulation pattern on comodulation masking release,"
J. Acoust. Soc. Am. 87, 269–283.

Hall, J.W., Grose, J.H. and Hatch,D.R. (**1996**):
"Effects of gating for signal detection in unmodulated and modulated noise,"
J. Acoust. Soc. Am. 100, 2365–2372.

Hellman, R.P. (**1972**):
"Asymmetry of masking between noise and tone,"
Percept. Psychophys. 11, 241–246.

v. Helmholtz, H.L.F. (**1863**):
Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik
F.Vieweg und Sohn, Braunschweig.

Hohmann, V., Kollmeier, B., and Müller-Deile, J. (**1997**):
"Festlegung der Parameter,"
in *Hörflächenskalierung – Grundlagen und Anwendung der kategorialen Lautheitsskalierung für Hördiagnostik und Hörgeräte-Versorgung*, edited by B. Kollmeier, (Buchreihe Audiologische Akustik, vol. 2, Median-Verlag, Heidelberg), Chap. 3.2, pp. 81–102.

Hübner, R., and Ellermeier, W. (**1993**):
"Additivity of loudness across critical bands: A critical test,"
Percept. Psychophys. 54, 185–189.

Jesteadt, W. (**1980**):
"An adaptive procedure for subjective judgements,"
Percept. Psychophys. 28, 85–88.

Kollmeier, B. (**1996**):
"Computer-controlled speech audiometric techniques for the assessment of hearing

loss and the evaluation of hearing aids,"
in *Psychoacoustics, Speech and Hearing Aids*, edited by B. Kollmeier (World Scientific, Singapore), 57–68.

Kumagai, M., Ebata, M., and Sone, T. (**1982a**):
"Comparison of loudness of impact sounds with and without steady duration (A study on the loudness of impact sound. II,"
J. Acoust. Soc. Jpn. (E) 3, 33–40.

Kumagai, M., Ebata, M., and Sone, T. (**1982b**):
"Loudness of impact sounds with wide-band spectrum (A study on the loudness of impact sound. III,"
J. Acoust. Soc. Jpn. (E) 3, 111–118.

Kumagai, M., Suzuki, Y., and Sone, T. (**1984**):
"A study on the time constant for an impulsive sound level meter (A study on the loudness of impact sounds. V,"
J. Acoust. Soc. Jpn. (E) 5, 31–36.

Kuwano, S., Namba, S., and Fastl., H. (**1988**):
"On the judgement of loudness, noisiness and annoyance with actual and artificial noises,"
J. Sound Vib., 457–465.

Kryter, K.D. (**1985**):
The effects of noise on man
Academic Press, London.

Langner, G., and Schreiner, C. (**1988**):
"Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanism,"
J. Neurophysiol. 60, 1799–1822.

Lawson, J. L., and Uhlenbeck, G. E. (**1950**):
Threshold Signals
(Radiation Laboratory Series, volume 24, McGraw Hill, New York.).

Levitt, H. (**1971**):
"Transformed up-down procedures in psychoacoustics,"
J. Acoust. Soc. Am. 49, 467–477.

Marks, L.E. (**1979**):
"Sensory and cognitive factors in judgements of loudness,"
J. Exp. Psychol.: Human Perception and Psychophysics 5, 426–443.

Marks, L.E. (**1978**):
"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.

Marks, L.E. (**1988**):
"Magnitude estimation and sensory matching,"
Percept. Psychophys. 43, 511–525.

Marks, L.E., and Warner, E. (**1991**):
"Slippery context effect and critical band,"
J. Exp. Psychol.: Human Perception and Psychophysics 17 , 986–996.

Marks, L.E. (**1994**):
"Recalibrating the auditory system: The perception of loudness,"
J. Exp. Psychol.: Human Perception and Psychophysics 20, 382–396.

McFadden, D.M. (**1986**):
"Comodulation masking release: Effects of varying the level, duration, and time delay of the cue band,"
J. Acoust. Soc. Am. 80, 1658–1667.

Mellert, V., and Reckhardt, C. (**1997**):
"Modellierung von Rangeeffekten bei der Bestimmung von Ispophonen,"
in Fortschritte der Akustik (pp. 498–499), DAGA 1997, Oldenburg, DEGA e.V..

Moore, B.C.J. (**1992**):
"Across-channel processes in auditory masking,"
J. Acoustic. Soc. Jpn. 13, 25–37.

Moore, B.C.J., and Alcantara, J.I., and Dau, T. (**1998**):
"Masking patterns for sinusoidal and narrowband noise maskers,"
submitted to J. Acoust. Soc. Am. .

Moore, B.C.J., and Glasberg, B.R. (**1983**):
"Suggested formulae for calculating auditory-filter bandwidths and excitation patterns,"
J. Acoust. Soc. Am. 74, 750–753.

Moore, B.C.J., Glasberg, B.R.,van der Heyden, M., and Kohlrausch, A. (**1995**):
"Comparison of auditory filter shapes obtained with notched–noise and noise–tone maskers,"
J. Acoust. Soc. Am. 97, 1175–1182.

Moore, B.C.J., and Jorasz, U. (**1991**):
"Detection of changes in modulation depth of a target sound in the presence of other modulated sounds,"
J. Acoust. Soc. Am. 91, 1051–1061.

Moore, B.C.J., and Jorasz, U. (**1996**):
"Modulation discrimination interference and comodulation masking release as a function of the number and spectral placement of narrow-band noise modulators,"
J. Acoust. Soc. Am. 100, 2373–2381.

Moore, B.C.J., Poon, P.W.F., Bacon, S.P., and Glasberg, B.R. (**1987**):
"The temporal course of masking and the auditory filters shape,"
J. Acoust. Soc. Am. 81, 1873–1880.

Müller, M., and Robertson, D. (**1991**):
"Relationship between tone burst discharge pattern and spontaneous firing rate of auditory nerve fibers in the guinea pig,"
Hear. Res57, 63–70.

Munson, W.A. (**1947**):
"The growth of auditory sensation,"
J. Acoust. Soc. Am. 19, 584–591.

Namba, S. (**1987**):
"On the psychological measurement of loudness, noisiness and annoyance: A Review,"
J. Acoust. Soc. Jpn. (E) 8, 211–222.

Namba, S., Hashimoto, T., and Rice, C.G. (**1987**):
"The loudness of decaying impulsive sounds,"
J. Sound Vib. 116, 491–507.

Niese, H. (**1959**):
"Die Trägheit der Lautstärke in Abhängigkeit vom Schallpegel,"
Hochfrequ. Elektroakust. 68, 143–152.

Niese, H. (**1965**):
"Beitrag zur Relation zwischen Lautstärke und Lästigkeit von Geräuschen,"
Acustica 15, 236–243.

Ogura, Y., Suzuki, Y., and Sone, T. (**1991**):
"A temporal integration model for loudness perception of repeated impulsive sounds,"
J. Acoust. Soc. Jpn. (E) 12, 1–11.

Ogura, Y., Suzuki, Y., and Sone, T. (**1993**):
"A new method for loudness evaluation of noises with impulsive components,"
Noise Control Engineering Journal 40, 231–240.

Ohm, G.S. (**1843**):
"Über die Defintion des Tones, nebst daran knüpfender Theorie der Sirene und ähnlicher Tonvorrichtungen,"
Ann. Phys. Chem. 59, 513.

Patterson, R. (**1976**):
"Auditory filter shapes derived with noise stimuli,"
J. Acoust. Soc. Am. 59, 640–654.

Patterson, R., and Nimmo-Smith, I. (**1980**):
"Off-frequency listening and auditory filter asymmetry,"
J. Acoust. Soc. Am. 67, 229–245.

Patterson, R., Nimmo-Smith, I., Holdsworth J., and Rice, P. (**1987**):
"An efficient auditory filterbank based on the gammatone function.,"
Meeting of the IOC Speech group on Auditory Modelling at RSRE (December 1987), 14-15..

Pedersen, O.J., Lyregaard, P.E., and Poulsen, T. (**1977**):
"The round robin test of impulsive noise,"
Report No. 22 (The Acoustics Laboratory, Technical University of Denmark, Lyngby).

Pollack, I. (**1964**):
"Neutralization of stimulus bias in auditory rating scale,"
J. Acoust. Soc. Am. 36, 1272–1276.

Port, E. (**1963a**):
"Über die Lautstärke einzelner kurzer Schallimpulse,"
Acustica 13, 212–223.

Port. E. (**1963b**):
"Zur Lautstärkeempfindung und Lautstärkemessung von pulsierenden Geräuschen,"
Acustica 13, 224–233.

Poulsen, T. (**1981**):
"Loudness of tones in a free field,"
J. Acoust. Soc. Am. 69, 1786–1790.

Poulton, E.C. (**1977**):
"Quantitative subjective assessments are almost biased, sometimes completely misleading,"
Br. J. Psychol. 68 , 409–425.

Poulton, E.C. (**1989**):
Bias in quantifying judgments
Lawrence Erlbaum, Hillscale, New Jersey.

Püschel, D. (**1988**):
Prinzipien der zeitlichen Analyse beim Hören
Ph.D. thesis, University of Göttingen.

Reckhardt, C., Mellert, V., and Kollmeier, B. (**1998**):
"Bestimmung von Isophonen mit einem Adaptiven Verfahren- Einfluß experimenteller Parameter auf die Ergebnisse,"
in Fortschritte der Akustik, DAGA 1998, Oldenburg, DEGA e.V., in press.

Reichardt, W. (**1965**):
"Zur Trägheit der Lautstärkebildung,"
Acustica 15, 345–354.

Reichardt, W. (**1970**):
"Subjective and objective measurement of loudness level of single and repeated impulses,"
J. Acoust. Soc. Am. 47, 1557–1562.

Reichardt, W., and Niese, H. (**1965**):
"Die Addition der Schallerregungen in den einzelnen Frequenzgruppen bei impulsiven Geräuschen,"
Acustica 16, 295–304.

Reichardt, W., and Niese, H. (**1970**):
"Choice of sound duration and silent intervals for test and comparison signals in the subjective measurement of loudness level,"
J. Acoust. Soc. Am. 47, 1083–1090.

Relkin, E.M., and Doucet, J.R, (**1997**):
"Is loudness simply proportional to the auditory nerve spike count,"
J. Acoust. Soc. Am. 101, 2735–2740.

Rule, S.J., and Curtis, D.W. (**1982**):
"Level of sensory and judgemental processing: Strategies for evaluation of a model,"
in B. Wegenrer (Ed.), *Social attitudes and psychophysical measurements*, Lawrence Erlbaum, Hillsdale, New Jersey.

Scharf, B. (**1959**):
"Loudness of complex sounds as a function of the number of components,"
J. Acoust. Soc. Am. 31, 783–785.

Scharf, B. (**1961**):
"Loudness summation and spectrum shape,"
J. Acoust. Soc. Am. 34, 228–233.

Scharf, B. (**1970**):
"Critical bands,"
in *Foundation of Modern Auditory Theory*, edited by J. V. (Academic, New York), Vol. 1..

Schneider, B. (**1988**):
"The additivity of loudness across critical bands: A conjoint measurement procedure,"
Percept. Psychophys 43, 211-222.

Schneider, B., and Parker, S. (**1990**):
"Does stimulus context affect loudness or only loudness judgment?,"
Percept. Psychophys 48, 409-418.

Schooneveldt, G.P. and Moore, B.C.J (**1987**):
"Comodulation masking release (CMR): Effects of signal frequency, flanking-band frequency, masker bandwidth, flanking-band level, and monotic versus dichotic presentation of flanking band,"
J. Acoust. Soc. Am. 82, 1944–1956.

Schooneveldt, G.P. and Moore, B.C.J (**1989a**):
"Comodulation masking release (CMR) as a function of masker bandwidth, modulator bandwidth, and signal duration,"
J. Acoust. Soc. Am. 85, 273–281.

Schooneveldt, G.P. and Moore, B.C.J (**1989b**):
"Comodulation masking release for various monaural and binaural combinations of the signal, on-frequency, and flanking bands,"
J. Acoust. Soc. Am. 85, 262–272.

Smith, R.L. (**1988**):
"Encoding of Sound intensity by auditory neurons,"
in *Auditory functions*, edited by G.M. Edelman, W.E. Gall, and W.M. Cowan (Wiley, New York), Chap. 8, pp. 243–274.

Smith, R.L., and Zwislocki, J.J. (**1975**):
"Short-term adaptation and incremental responses of single auditory-nerve fibers,"
Biol. Cybern., 169–182.

Sone, T., Suzuki, Y., Kumagai, M., and Takahashi, T. (**1986**):
"Loudness of a single burst of impact sound: Results of round robin test in Japan (I),"
J. Acoust. Jpn. (E) 7, 173–182.

Stevens, S.S., and Galanter, E.H. (**1957**):
"Ratio scales and category scales for a dozen perceptual continua,"
J. Exp. Psychol. 54, 377–411.

Stone, M.A., Moore, B.C.J., and Glasberg, B.R. (**1997**):
"A real-time DSP-Based loudness Meter,"
in 7th Oldenburg Symposium on Psychoacoustics, ed. A Schick, M Klatte, BIS , Oldenburg, 1997, p. 587–602.

Strickland, E. A. and Viemeister, N.F. (**1997**):
"The effect of frequency region and bandwidth on the temporal modulation transfer function,"
J. Acoust. Soc. Am. 102, 1799–1810.

Tachibana, H., Ishizaki, S., and Yoshihisa, K. (**1987**):
"A method of evaluating the loudness of isolated impulsive sounds with narrow frequency components,"
J. Acoust. Soc. Jpn. (E) 8, 29–38.

Takeshima, H., Suzuki, Y., Kono, S.,and Sone, T. (**1988**):
"Growth of the loudness of a tone burst with duration up to 10 seconds,"
J. Acoust. Soc. Jpn. (E) 9, 295–300.

Verhey, J.L., and van de Par, S. (**1997**):
"Messungen und Modellrechnungen zu CMR,"
Fortschritte der Akustik - DAGA'97, DEGA, Oldenburg.

Verhey, J.L., Dau, T., and Kollmeier, B. (**1998**):
"Within-channel cues in comodulation masking release (CMR): Experiments and model predictions using a modulation-filterbank model,"
submitted to the J. Acoust. Soc. Am. .

Viemeister, N.F. (**1979**):
"Temporal modulation transfer functions based upon modulation thresholds,"
J. Acoust. Soc. Am. 66, 1364–1380.

Zeng, F.-G., and Shannon, R.V. (**1991**):
"Loudness balance between electric and acoustic stimulation,"
Science 264, 564–566.

Zwicker, E. (**1956**):
"Die elementaren Grundlagen zur Bestimmung der Informationskapzität des Gehörs,"
Acustica 6, 365–381.

Zwicker, E. (**1965**):
"Temporal effects in simultaneous masking and loudness,"
J. Acoust. Soc. Am. 37, 132–141.

Zwicker, E. (**1966a**):
"Ein Beitrag zur Lautstärkemessung impulshaltiger Schalle,"
Acustica 17, 11–22.

Zwicker, E. (**1966b**):
"Ein Beitrag zur Unterscheidung von Lautstärke und Lästigkeit,"
Acustica 17, 22–25.

Zwicker, E. (**1969**):
"Der Einfluß der zeitlichen Struktur von Tönen auf die Addition von Teillautheiten,"
Acustica 21, 16–25.

Zwicker, E. (**1974**):
"Loudness and excitation patterns of strongly frequency modulated tones,"
in Sensation and Measurement, Paper in honor of S. S. Stevens, (D. Reidel, Dordrecht/Boston), 325–335.

Zwicker, E. (**1977**):
"Procedure of calculating loudness of temporally variable sounds,"
J. Acoust. Soc. Am. 62, 675–682.

Zwicker, E. (**1984**):
"Dependence of post-masking on the masker duration and its relation to temporal effects in loudness,"
J. Acoust. Soc. Am. 75, 219–223.

Zwicker, E. and Fastl, H. (**1972**):
"On the development of the critical band,"
J. Acoust. Soc. Am. 52, 699-702.

Zwicker, E. and Fastl, H. (**1990**):
Psychoacoustics
(Springer-Verlag, Berlin Heidelberg, 1990).

Zwicker, E., Flottorp, G., and Stevens, S.S. (**1957**):
"Critical bandwidth and loudness summation,"
J. Acoust. Soc. Am. 29, 548–557.

Zwicker, E., and Scharf, B. (**1965**):
"A model of loudness summation,"
Psychol. Rev. 72, 3–26.

Zwicker, E. and Schorn, K. (**1982**):
"Temporal modulation transfer functions based upon modulation thresholds,"
Audiology 21, 474–492.