# Chapter 7
# Where Mathematics and Hearing Science Meet: Low Peak Factor Signals and Their Role in Hearing Research

**Armin Kohlrausch and Steven van de Par**

**Abstract**  In his scientific work, Manfred Schroeder touched many different areas within acoustics. Two disciplines repeatedly show up when his contributions are characterized: his strong interest in mathematics and his interest in the perceptual side of acoustics. In this chapter, we focus on the latter. We will first give a compressed account of Schroeder's direct contributions to psychoacoustics, and emphasize the relation with other acoustics disciplines like speech processing and room acoustics. In the main part of the chapter we will then describe psychoacoustic work being based on or inspired by ideas from Manfred Schroeder. Due to Schroeder's success in securing a modern online computer for the Drittes Physikalisches Institut after returning to Göttingen in 1969, his research students had a head start in using digital signal processing in room acoustics for digital sound field synthesis and in introducing digital computers into experimental and theoretical hearing research. Since then, the freedom to construct and use specific acoustic stimuli in behavioral and also physiological research has grown steadily, making it possible to test many of Schroeder's early ideas in behavioral experiments and applications. In parallel, computer models of auditory perception allowed users to analyze and predict how specific properties of acoustic stimuli influence the perception of a listener. As in

A. Kohlrausch (✉)
Behavior, Cognition, & Perception Group, Philips Research Europe, High Tech Campus 36 (MS WO 02), Eindhoven 5656AE, The Netherlands

Human-Technology Interaction, Eindhoven University of Technology, Eindhoven, The Netherlands
e-mail: armin.kohlrausch@philips.com

other fields of physics, the close interplay between experimental tests and quantitative models has been shown to be essential in advancing our understanding of human hearing.

## 7.1 Introduction

One of the scientific areas that was close to the heart of Manfred Schroeder was psychoacoustics. During his first period in Göttingen, the focus of his work was on the physical and statistical side of acoustics. On moving to the Bell Laboratories in 1954, he came into an environment with a long history in hearing and speech research, started by Harvey Fletcher in the 1910s. In his autobiographic chapter, Schroeder describes a number of examples how he got interested in the perceptual side of acoustics, both from his research in room acoustics and in speech processing. During his time as professor at the University of Göttingen, the number of diploma and Ph.D. students focusing on hearing related studies increased continuously. This increased interest was supported by a rebuilding of the central space in the "Halle," in front of the "Reflexionsarmer Raum," where the control panels for the loudspeaker dome (well known to acousticians from a photo in Blauert's book on spatial hearing, see Fig. 3.50 in [2]) were dismantled and spaces for two listening booths were created—unofficially "owned" by the two research groups around Birger Kollmeier and the first author. In parallel with the acoustic spaces, the computer infrastructure for controlling listening experiments and generating acoustic stimuli also grew steadily. This experimental infrastructure and a growing group of young scientists were essential for the increasing level of sophistication of hearing research at the Drittes Physikalisches Institut (DPI).

In this chapter, we want to briefly summarize Schroeder's contributions to the psychoacoustic literature. In the major part of this chapter, we will describe how the particularly close link between hearing research and another of Manfred Schroeder's scientific interests, defining signals with very specific properties, like low peak factors, has influenced the psychoacoustic community. Some aspects of these developments are also closely interlinked with other facets of Schroeder's life, as will become clear by comparing the present text with the corresponding passages from his autobiographic chapter.

The advanced digital signal processing capabilities at the DPI, quite unusual for an institute of physics in the 1970s and strongly inspired by Schroeder's experience at the Bell Laboratories, allowed a high level of creativity in constructing signals with specific spectral and temporal properties in both listening experiments and model simulations. These innovative approaches, which started at the DPI, spread to other places like Eindhoven and Oldenburg, and later on to Lyngby/Kopenhagen, and have influenced research paradigms in hearing research groups all over the world.

## 7.2     Schroeder's Major Contributions to Hearing Research

Schroeder's interest in psychoacoustics had a number of roots. Through his work on efficient speech coding and the naturalness of synthesized speech, he developed an interest in the role of the signal envelope and fine structure on timbre. A specific manifestation was his work on "monaural phase effects." Furthermore, he used concepts of perceptual masking and perceptual distance measures in speech coding algorithms to improve the trade-off between perceptual quality and bit rate. Through his work on concert hall acoustics and digital reverberators, he recognized the great importance of spatial parameters such as the interaural cross-correlation, for good perceptual quality of a reverberant environment; this had led to work on binaural modeling. Actually, stimulated by his research in concert hall acoustics in the 1970s, binaural psychoacoustics evolved in the 1980s into one of the strongholds of his research group at the DPI in Göttingen. And finally, he had a genuine interest in developing and improving models of the auditory periphery—models that could help to understand specific perceptual phenomena, like the level dependence of difference tones. An excellent overview about how deeply he thought about these various aspects of hearing science can be found in his overview article from 1975: Models of Hearing [59].

Probably the psychoacoustic topic that interested Schroeder the earliest was the question of the extent to which the human hearing system could decode the phase spectrum of a signal. His first publication on this topic was an abstract from the 58th meeting, in 1959, of the Acoustical Society of America with the title: New results concerning monaural phase sensitivity [54]. Schroeder describes the perceptual consequences of changing the component phase in harmonic complex signals comprising up to 31 components. Aspects mentioned are the strong influence of the peak factor on the signal timbre, the absence of timbre changes for signals with identical envelopes, and the possibility to create strong and varying pitch phenomena allowing one to play melodies, just by changing phases of individual components.

The emphasis on the relation between timbre and waveform had a direct link to the work on vocoder quality, where, on the synthesis side, the excitation signal for voiced speech was composed of harmonic complex tones for which one could choose the relative phases freely. Schroeder and colleagues had observed that the excitation function strongly influenced the quality of vocoder speech (see, e.g., [55]). Schroeder described this early work, including waveform examples for different phase choices, in his 1975 IEEE paper [59] where he also referred to an earlier, less-known account by R. L. Pierce on this work in the American Scientist from 1960 [50]. In the course of his research on monaural phase effects, he also formulated closed solutions for generating periodic signals with low peak factors [57]. This short mathematical paper from 1970 remained relatively unknown for nearly 20 years, receiving only 29 citations until 1989. After this phase rule was introduced into hearing research in 1986 in [69] (see next section), the number of citations exploded and grew to a total of 312 (as of October 2013), making it,

together with the paper on measuring reverberation time, by far the most cited of his scientific publications.

Through his work on efficient speech coding, eventually resulting in code-excited linear prediction, Schroeder became acquainted with the phenomenon of masking and the great value of perceptually based distance metrics (see also the Chap. 10 in this book). The term "subjective error criteria" appeared first in the title of a conference contribution by Atal and Schroeder in 1978 [1]. As stated in the abstract of that paper, the human ear does not use simple RMS error measures when judging distortions introduced through coding. The new approach by Atal and Schroeder was to minimize the subjective distortion by shaping the spectrum of the resulting prediction error to be optimally masked by the speech spectrum. This approach does not reduce the total power of the error signal, which is determined by the quantizer, but redistributes its power along the frequency axis to have a more constant signal-to-noise ratio at all frequencies.

The concept of masking and its application in speech coding appeared in a follow-up article in 1979, jointly written with Atal and J.L. Hall [62]. The paper contains an extensive description of how to transform a short-term power spectrum into an excitation pattern, including transformation through outer and middle ear, calculation of critical-band densities, and transforming these into an excitation pattern from which the loudness was finally computed. The masked thresholds due to a masking signal, which in this application was the speech signal, were computed by multiplying the signal (speech) excitation function by a sensitivity function. In this way, the authors were able to predict whether, for a given short-term spectrum of the speech signal, a specific noise would fall below this masking function and was therefore inaudible. In addition to predicting the masked threshold, they derived, for suprathreshold noise levels, an objective degradation scale by relating the speech and the noise loudness (which, in modern terms, was calculated as partial noise loudness in the presence of the speech signal). This approach, in which state-of-the-art perception science was functionally integrated in a real-life signal-processing application, was later on also successful in the development of perceptual audio coding starting in the mid-1980s. An early account of this audio coding work, referring back to the earlier work by Schroeder et al. [62], is the paper by Johnston: Transform coding of audio signals using perceptual noise criteria [28].

An early account of spatial perceptual aspects in concert hall acoustics is given in the context of the acoustic measurements in the Philharmonic Hall [63]. One of the analyzed parameters which varied strongly between listener positions was the directional distribution of the early energy within 50 ms of the direct sound. After the acoustic changes in the hall had been finished, the variability of this parameter was much reduced. In their conclusion, the authors state, "It is therefore tentatively concluded that the directional distribution of early reflections is a significant contributing factor to acoustical quality" (p. 440 in [63]). This recognition of the role of spatial dissimilarity for good concert hall acoustics was further supported by the large comparative study of European concert halls, performed together with Gottlob and Siebrasse in Göttingen [64]. From their multidimensional analysis of

paired comparison data from many listeners, the interaural coherence turned out to be an independent dimension, being negatively correlated with subjective preference. This article addresses spatial hearing also from a different perspective. It describes the setup in the anechoic chamber in Göttingen, which permitted the reproduction of two-channel binaural recordings via two loudspeakers in an anechoic environment. The critical element in the reproduction chain was a cross talk cancellation computation, as it had been described and demonstrated 10 years earlier by Schroeder and Atal [61]. With cross talk cancellation the sound transmission from two loudspeakers could be directed to the two ears as otherwise only possible by headphone reproduction.

In his paper "New viewpoints in binaural interaction" at the fourth International Symposium on Hearing in Keele, 1977 [60], Schroeder made an original contribution to theories of binaural hearing, inspired by his work on basilar membrane characteristics. The core of his proposal used cochlear delays as a basis for the analysis of interaural delays. The computation of waveform delays between right and left ear would, instead of being enabled by interaural neural delay lines (as they are included in many types of binaural models), be realized by comparing interaurally the activities between different places on each basilar membrane. Such a comparison was possible and meaningful, because the basilar membrane activity has, for a given signal, a systematic relation between place and delay, and the *basilar membrane delay* values are of the magnitude necessary to analyze realistic *interaural delay* values. This proposal was later implemented and evaluated by Shamma and colleagues [67] who coined the term "stereausis" for this scheme. A consequence of this way of thinking about interaural delay, as pointed out by Joris and colleagues at the 13th International Symposium on Hearing in Dourdan, France [29], was that the traditional delay-line model using neural delay lines needed to have a very exact anatomical link between equivalent basilar membrane positions in the two cochleae. Only small spatial mismatches would introduce considerable offsets in the interaural delay values, particularly at low frequencies.

A final topic which we want to mention is Schroeder's research on basilar membrane modeling and its relation to specific psychoacoustic phenomena. Schroeder's interest in the relation between basilar membrane and hair cell properties on one side and perceptual observations on the other might have its roots in his attempts to relate cubic difference tone (CDT) data to the critical-band concept and basilar membrane mechanics. In his first paper from 1969 [56], he analyzed CDT data from Goldstein, and added basilar membrane simulation and additional CDT phase measurements performed at Bell Laboratories. As in similar earlier work by Zwicker, this established a close relation between CDT generation and auditory excitation. In addition, Schroeder concluded that a "purely mechanical model of the ear" was insufficient to explain the amplitude and phase characteristics of the $2f_1 - f_2$ difference tone. An additional amplitude nonlinearity was required, and Schroeder concluded, based on the sign of the resulting CDT, that it had to be an amplitude-*limiting* nonlinearity.

In the following years, Schroeder, often together with J.L. Hall, contributed papers on basilar membrane models [58] and a model for mechanical to neural transduction in the inner hair cell [65], both of which have been highly influential in the development of more realistic models for the auditory periphery. How useful fast and realistic time-domain basilar membrane models can be for the interpretation of psychoacoustics results will be demonstrated in the next section.

## 7.3 Harmonic Complex Tone Stimuli and the Role of Phase Spectra

A particular class of auditory stimuli are signals with a periodic waveform. They can be analyzed in terms of their temporal properties (e.g., the perceptual influence of a slight deviation from temporal regularity in an otherwise regular series of clicks, see [43]) or in the context of their spectral properties (consider, for example, the role of the vocal tract filter on the resulting vowel quality). Of course, from a mathematical point of view, time domain and spectral descriptions are fully equivalent, as long as the complex spectrum, including phase, is considered. Historically, however, temporal and spectral views were quite distinct, mainly because of the influence of signal analysis systems that represented the power spectrum but not the phase spectrum. Also, the psychoacoustic paradigm of critical bands and auditory filters, which for a long time were only defined in terms of their overall bandwidth and their amplitude characteristics (see, e.g., [19]), made it difficult to bring the temporal and spectral views closer together. Again, the increasing use of computer programs to generate acoustic stimuli and to perform time-domain modeling of perceptual processes emphasized the role of the phase spectrum on the perceptual quality of periodic signals (e.g., [16]). In addition, for us students in Göttingen, there was the strong interest of Schroeder in phase effects which prepared us more than other colleagues at that time to think in terms of signal waveforms, envelopes, and the time dimension (e.g., [66]). In the following, we will focus on one specific type of complex tones, tones with so-called Schroeder phases.

### 7.3.1 Schroeder-Phase Harmonic Complex Tones

#### 7.3.1.1 Definition

The term "Schroeder phase" refers to a short paper by Schroeder from 1970 [57]. In this paper, he addressed the problem how the peak-to-peak amplitude of a periodic waveform can be minimized. His proposed solution lies in a phase choice which gives the signal an FM-like property. The general solution derives the individual phase values without any restriction on the amplitude spectrum. A specific case is

that of a harmonic complex with a flat power spectrum having a band-pass characteristic. The solution for the individual phase values of such a complex is as follows:

$$\phi_n = -\pi n(n-1)/N, \tag{7.1}$$

with $N$ the total number of components in the complex. The important term in this equation is the quadratic relation between component number, $n$, and component phase, $\phi_n$, which leads to an approximately linear increase in instantaneous frequency over time. The normalization with $N$ creates a signal, for which the instantaneous frequency sweeps once per period from the frequency of the lowest to that of the highest component in the complex. In fact, the instantaneous frequency has a periodic sawtooth-like course for such Schroeder-phase signals.

It is obvious that reversing the initial sign in (7.1) has no influence on the peak factor of the resulting signal, but it will invert the direction of the linear frequency sweep. Because these two versions of a Schroeder-phase signal lead to substantially different percepts, a convention has been introduced to distinguish them. A *negative* Schroeder-phase signal is a signal were the phases of individual components are chosen as in (7.1). In contrast, a *positive* Schroeder-phase signal has phase values with a positive sign in front of the fraction. One can memorize this relation by using the fact that the sign of the phase is opposite to the direction of change in instantaneous frequency.

### 7.3.1.2   Acoustic Properties

By construction, Schroeder-phase stimuli have a relatively flat temporal envelope and the peak factor, defined as the ratio between envelope maximum and the RMS-value of the signal, is much lower than for other phase choices. This is demonstrated in Fig. 7.1, which compares the waveforms for harmonic complexes composed of 19 equal-amplitude harmonics of a 100-Hz fundamental from 200 to 2,000 Hz. There are three different choices of the component phases: positive Schroeder phase, negative Schroeder phase, and, in the top part, a zero-phase stimulus. For this latter stimulus, the energy is concentrated at very short instances within each period, leading to a much higher peak factor. The spectro-temporal properties of these signals can be seen more clearly in a short-time spectral representation.

Figure 7.2 shows the spectra of the three signals from Fig. 7.1 calculated using a moving 5-ms Hanning window. The sawtooth-like frequency modulation of the two Schroeder-phase complexes is pronounced in this representation. In addition, the plot for the zero-phase complex shows ridges at the spectral edges of 200 and 2,000 Hz. These relative spectral maxima can be perceived as pitch, superimposed on the 100-Hz virtual pitch of the complexes, and the presence of these pitch percepts has been used as a measure of the internal representation of such harmonic complexes [34, 35]. The visibility of the temporal structures in the short-time

**Fig. 7.1** Time functions of harmonic complexes for three different choices of the component's starting phases. *Top*: $\phi_n = 0$ zero-phase complex, *middle*: $\phi_n = -\pi n(n-1)/N$, negative Schroeder-phase complex, *bottom*: $\phi_n = +\pi n(n-1)/N$, positive Schroeder-phase complex. All complexes are composed of the equal-amplitude harmonics 2 to 20 of fundamental frequency 100 Hz. For this plot, the amplitude of an individual harmonic was set at 1. Reprinted from [36]. Copyright (1995) Acoustical Society of America
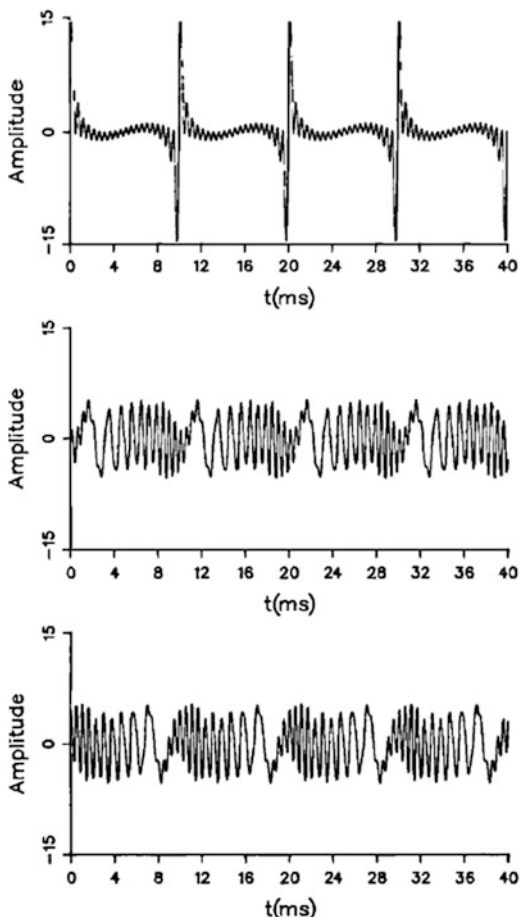
spectrum depends critically on the duration of the temporal window, relative to the period of the sound. The shorter the window, the more visible are temporal changes; the longer the window, the more the spectral properties are emphasized.

## 7.3.2 Role in Hearing Research and Perceptual Insights

The first paper in which the Schroeder-phase formula was used in hearing experiments was published by Mehrgardt and Schroeder in the proceedings of the 6th International Symposium on Hearing, 1983 [41]. In this paper, the quadratic phase formula from (7.1) was combined with an additional scaling factor, which permitted control of the spread of signal energy throughout the period. The spectrum of the harmonic complex was, however, not flat as in most later
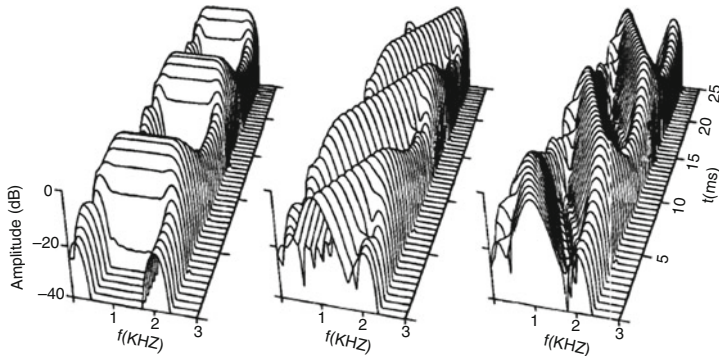
**Fig. 7.2** Short-time spectral representation of the signals from Fig. 7.1 using a Hanning window with length 5 ms. The *left* panel shows the zero-phase signal, the *middle* panel the negative Schroeder phase and the *right* panel the positive Schroeder-phase complex. Reprinted from [36]. Copyright (1995) Acoustical Society of America

investigations, but the individual components had Hanning-weighted amplitudes. This paper emphasized the influence of the masker's temporal waveform on the observed masking behavior and showed how strongly the acoustic waveform and the resulting masked thresholds can vary by just varying the phase spectrum.

The great potential of Schroeder-phase signals to expose the phase characteristic of an auditory filter was discovered quite accidentally. During his master thesis research, Bennett Smith, traveling back and forth between Göttingen and Paris, where he continued to work as an engineer at the IRCAM, was interested in acoustic figure-ground phenomena, translating spatial orientation in the visual domain into linear frequency modulation in the auditory domain [68]. In the construction of his acoustic background stimuli, he made use of the Schroeder-phase formula. He did, unfortunately, not find any effect of acoustic figure-ground orientation on audibility, but made instead another observation: The audibility of a sinusoidal stimulus in a complex-tone masker depended strongly on the sign in the phase formula: When the background was constructed with a positive sign, the masked thresholds were lower by up to 20 dB, compared to the situation in which the phase sign was negative.

This large threshold difference formed a considerable scientific puzzle which kept the participants of Schroeder's weekly seminar busy for a prolonged period: Both masker versions had a similarly flat temporal envelope, so there was no reason to assume a difference in masking potential for simultaneously presented sinusoids. In addition, no temporal asymmetry, as in backward versus forward masking, could play a role given the simultaneous presentation of masker and test signal. We even considered the effects of FM to AM conversion of the masker waveform that should occur when the linear frequency modulation of the masker interacts with the shallow slopes of the auditory filter, without arriving at a satisfactory explanation. The effect was finally understood on the basis of computer simulations with a time-domain basilar membrane model which had been realized by Hans-Werner Strube

[70]. This model was, at that time, the first computer model with a realistic phase characteristic for the basilar membrane, which was able to compute the basilar membrane output in the time domain while being sufficiently time efficient. Strube observed that the two versions of the Schroeder-phase complex led to very different waveforms at the output of the various model elements that each represented a specific place on the basilar membrane. One waveform was clearly more modulated, and the existence of valleys in the masker waveform fitted nicely with the observation by Smith that this masker also led to lower masked thresholds: The target signal could be detected easier, i.e., at a lower level, within these masker valleys—a phenomenon for which the term "listening in the valleys" is used in the psychoacoustic literature. On the basis of the consistent explanation made possible by these simulations, we dared to summarize our experimental data and submit the manuscript to JASA. For me (AK), this was to become my first internationally peer-reviewed publication, and actually the only one jointly co-authored by Manfred Schroeder [69].

In the next years, a great many further experiments and model simulations were performed in Göttingen [30, 36], which led to the following insights: The clue to understanding the differences between positive and negative Schroeder-phase stimuli lies in the phase characteristic of the auditory filter, in combination with some general rules in masking of narrowband signals. If subjects have to detect a narrowband stimulus in a broadband masker, only the frequency region around the target frequency is of interest. In order to simulate the transformations in the auditory periphery for such an experiment, we have to compute the waveforms of the acoustic stimuli at the output of the auditory filter that is centered on the target frequency. For random noise maskers, only the amplitude characteristic of the filter is relevant, in line with the early "critical-band" concepts. For periodic signals as considered here, however, this filtered waveform, and its masking behavior, will depend critically on the amplitude and the *phase* characteristic of the auditory filter.

The most important conclusion was that, for the right choice of stimulus parameters, the phase characteristic of a Schroeder-phase stimulus matches quite closely the phase characteristic of the auditory filter, at least in the spectral region of maximum transfer of the filter. Since Schroeder-phase stimuli come in two flavors, one version, the negative Schroeder-phase stimulus, will have the same phase characteristic as the auditory filter, while the positive Schroeder-phase stimulus has a phase spectrum that is *opposite* to that of the filter. In the transfer through a filter, the input phase spectrum and the filter phase spectrum add to give the phase spectrum of the output signal. For the positive Schroeder phase, we thus have the situation of phase compensation, and the resulting filtered signal at the output of the auditory filter has a nearly constant phase of the strongest components, those with a frequency at the center of the corresponding basilar membrane filter, resulting in a highly peaked signal. Conceptually, this situation is quite similar to pulse compression through frequency modulation, as it is used in radar and sonar technology. In a way, the positive Schroeder-phase stimuli are matched in their phase characteristic to the auditory filters as they are realized mechanically in the inner ear.

**Fig. 7.3** Responses of a linear basilar membrane model at resonance frequency 1,100 Hz to the three harmonic maskers with fundamental frequency 100 Hz shown in Fig. 7.1. The *top* panel shows the zero-phase signal, the *middle* panel the negative Schroeder phase and the *bottom* panel the positive Schroeder-phase complex. Reprinted from [36]. Copyright (1995) Acoustical Society of America
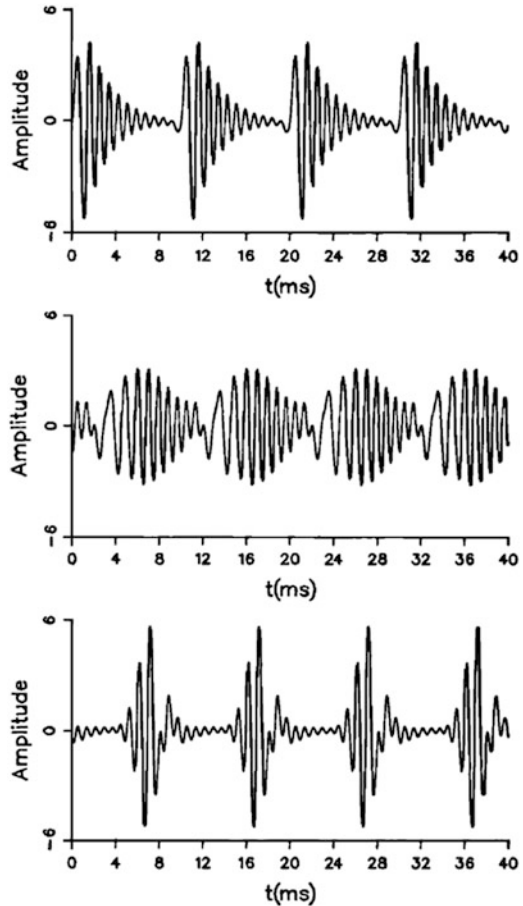
This interpretation leads to some interesting, counterintuitive predictions: If a specific Schroeder-phase stimulus is optimally matched in its phase to the inner-ear filter at a certain frequency, then such a signal should have a higher peak-factor after filtering than a zero-phase input stimulus. This prediction is analyzed in Fig. 7.3, which shows waveforms of the three stimuli from Fig. 7.1 at the output of a linear basilar membrane filter tuned to 1,100 Hz. The two panels at the bottom show the waveforms of the two Schroeder-phase complexes. While the broadband input signals to the filter have a flat temporal envelope (cf. Fig. 7.1), both waveforms have a clear amplitude modulation after filtering, which follows the 10-ms periodicity of the stimulus. But the depth of modulation is quite different for the two signals, the positive Schroeder-phase stimulus at the bottom has a much higher peak factor and a longer period of small envelope values than the negative Schroeder-phase complex in the middle. This simulation reflects the initial observation made by Strube in [70].
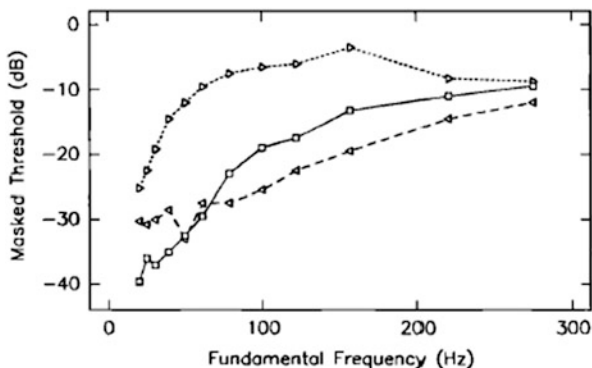
**Fig. 7.4** Simultaneous masked thresholds of a 260-ms, 1,100-Hz signal as a function of the fundamental frequency $f_0$ of the harmonic complex masker. Thresholds are expressed relative to the level of a single masker component. The maskers were presented at a level of 75 dB SPL. *Squares*: zero-phase complex; *right-pointing triangles*: negative Schroeder-phase complex; *left-pointing triangles*: positive Schroeder-phase complex. Reprinted from [36]. Copyright (1995) Acoustical Society of America

The top panel shows the filtered version of the zero-phase complex, which has a highly peaked input waveform. The filtered waveform reflects, within each period, the impulse response of the auditory filter, because the zero-phase complex is similar to a periodic sequence of pulses. By careful analysis one can even see some properties of the filter's frequency-dependent group delay: around each peak, low-frequency components with a long waveform period occur first, followed later on by higher-frequency components. Comparing the top and the bottom panel reveals that, indeed, the positive Schroeder-phase complex has a somewhat higher peak factor than the zero-phase complex, and its energy is more concentrated in time, as expected based on the pulse-compression analogon.

The relation between zero-phase and positive-Schroeder-phase stimuli also formed the key to estimating the phase properties of a specific point on the basilar membrane. If we were able to determine the phase curvature, for which the match between stimulus and filter phase is "optimal," then this value would be an indication of the phase curvature (or the frequency-dependent group delay) of the filter. The clue to such an analysis is given by comparing perceptual thresholds for positive Schroeder-phase maskers with those for zero-phase maskers. For maskers, for which this difference is largest (and for which the Schroeder-phase stimulus as masker gives lower masked thresholds), the phase curvature at the signal frequency is an estimate of the filter phase. Figure 7.4 shows data from [36] which were used for such a computation.

In the region of $f_0$ values between 100 and 150 Hz, the differences in thresholds obtained with the positive Schroeder-phase complex (left-pointing triangles) and with the zero-phase complex (squares) are largest. For theses complexes, the second derivative of the phase-versus-frequency relation, which indicates the phase curvature has values between $1.05 \times 10^{-5}\pi/\text{Hz}^2$ and $0.74 \times 10^{-5}\pi/\text{Hz}^2$. We can

conclude that the curvature of the phase characteristic for the basilar membrane filter centered at 1,100 Hz should be in the range of these two values. A similar conclusion about the phase curvature can be derived from the parameters of those complexes, for which positive Schroeder-phase and zero-phase complexes lead to approximately the same threshold. In this case, the internal envelope modulation of the two complexes after filtering on the basilar membrane should be approximately equal. As explained in detail in [36], the phase curvature in the Schroeder-phase stimuli should be half the value of the filter curvature, and this is reached for fundamental frequencies of 50–75 Hz. And exactly in this region, the two lower curves in Fig. 7.4 cross each other.

This consideration allowed a first computation of the auditory filter phase for one frequency, 1,100 Hz. In [36], additional threshold measurements were included for frequencies 550, 2,200 and 4,400 Hz, thus covering a range of 3 octaves. It is often assumed that the auditory filter has a nearly constant quality factor across the range of audible frequencies. If this scaling relation were to hold also for the phase characteristic, then the results obtained at 1,100 Hz would allow a direct prediction for the phase characteristic in the range of 3 octaves around 1,100 Hz. The comparison with the results at 550 Hz indeed revealed the expected relation, while towards higher frequencies, the curvature changed somewhat less with center frequency than expected for a system in which the amplitude and phase characteristics of the filters remain constant on a logarithmic frequency scale.

One final important observation from these initial Schroeder-phase studies needs to be mentioned. In the 1980s, the view on the shape of auditory filters was strongly influenced by the work of Patterson, Moore and colleagues, who had used the notched-noise technique to estimate the amplitude characteristic of the auditory filter. The best characterization was possible with a so-called rounded exponential filter shape (see, e.g., [49]). A time-domain implementation of a filter with such an amplitude characteristic was possible based on so-called gamma-tone filters [27, 48]. Because of the many studies supporting this concept of auditory filters, we were interested to analyze the Schroeder-phase stimuli with such a filter.

Figure 7.5 presents, in a format similar to Figs. 7.1 and 7.3, four periods of the waveform for harmonic complexes with fundamental frequency 100 Hz. The analysis shows the output of the gamma-tone filter tuned to 1,100 Hz, and the three subpanels are for the three different phase choices. It is apparent that this filter does not lead to differences in the modulation depth between the three stimuli, and based on this simulation one would expect quite similar masking behavior of all three complex tones, in contrast to the experimental data. The major reason for the similar treatment of the two Schroeder-phase maskers by the gamma-tone filter is its antisymmetric phase characteristic close to its resonance frequency. The curvature of the filter phase changes its sign at the resonance frequency from negative to positive. A filter with such a phase characteristic can never flatten out the phase of a Schroeder-phase complex that changes over the full range of its passband, and should therefore be used with caution in experiments where signal phase matters. As has been shown later by Lentz and Leek [40], this shortcoming of

**Fig. 7.5** Responses of a
linear, fourth-order gamma-
tone model at resonance
frequency 1,100 Hz to the
three harmonic maskers
with fundamental frequency
100 Hz shown in Fig. 7.1.
The *top* panel shows the
zero-phase signal, the
*middle* panel the negative
Schroeder phase and the
*bottom* panel the positive
Schroeder-phase complex.
Reprinted from
[36]. Copyright (1995)
Acoustical Society of
America
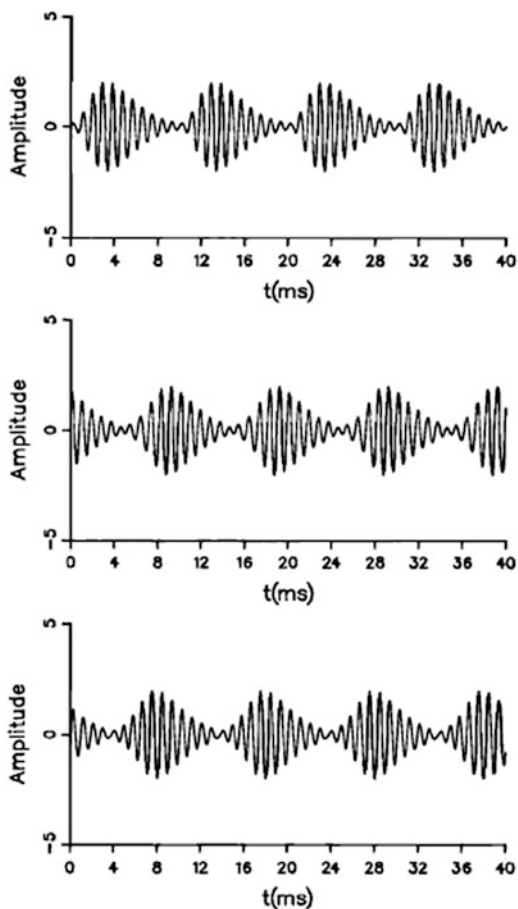


the gamma-tone filter can be overcome by using a nonlinear extension of this filter,
the gammachirp filter proposed by Irino and Patterson [26].

### 7.3.3  Later Developments

Although the first paper on Schroeder-phase stimuli was already published in 1986
([69]), the paradigm was only widely adopted after publication of our second paper
in 1995 [36]. The first papers that used the term "Schroeder phase" in their title were
published in 1997 [9, 10]. Many authors related psychoacoustic findings with
Schroeder-phase stimuli to the properties of the basilar membrane. Differences
that were found between normal-hearing and hearing-impaired subjects and also
influences of the overall presentation level indicated some role of active processes

in creating large differences between positive and negative Schroeder-phase stimuli [10, 71, 72]. Based on the results of these studies, Summers concluded: "The current results showed large differences in the effectiveness of positive and negative Schroeder-phase maskers under test conditions associated with nonlinear cochlear processing. The two maskers were more nearly equal in effectiveness for conditions associated with more linear processing (high levels, hearing-impaired listeners). A number of factors linked to the cochlear amplifier, including possible suppressive effects and level-dependent changes in the phase and magnitude response of effective filtering, may have contributed to these differences." [[71], p. 2316].

The analysis of the phase characteristics of auditory filters was further refined by Oxenham and Dau [44, 45]. They varied the phase curvature of Schroeder-phase complexes by using a scalar multiplier in front of (7.1), very similar to the use of the Schroeder-phase formula by [41]. They concluded that the scaling invariance of filter phase with filter center-frequency, as expected for a set of filters with constant quality factor, might hold for frequencies above 1 kHz, but not for lower frequencies.

Schroeder-phase stimuli have also been used in physiological experiments, which allowed a direct test of the basic hypothesis of the role of peripheral filtering on modulation depth of the waveform, as published in 1985 by Strube. In 2000, Recio and Rhode [52] measured the basilar membrane response in the chinchilla for positive and negative Schroeder-phase stimuli and also for clicks, thus using types of stimuli very similar to the early psychoacoustic studies. They concluded: "The behavior of BM responses to positive and negative Schroeder complexes is consistent with the theoretical analysis performed by Kohlrausch and Sander in 1995, in which the curvature, i.e., the second derivative of the phase versus frequency curve of the BM was used to account for the differences in the response to each of the two Schroeder phases... Hence, phase characteristics of basilar membrane responses to positive Schroeder-phase stimuli show reduced curvatures (relative to the stimulus), and, as a result, peaked waveforms (Kohlrausch and Sander, 1995)" [[52], p. 2296].

Given the relevance of signal phase in Schroeder's work and in hearing research, the first author chose this topic in his contribution to the session "Honoring Manfred R. Schroeder, his contributions and his life in acoustics" at the 4th joint meeting of the Acoustical Society of America and the Acoustical Society of Japan in Honolulu, Hawaii (2006) in his talk: "Schroeder's phase in psychoacoustics." This session was held in the period that the DPI building at the address Bürgerstr. 42–44 was being dismantled and rebuilt for non-academic use. In order to preserve some memories, I (AK) was able to physically remove and to hand over to Schroeder a sign from the institute's parking lot, indicating the place where he and his family used to park their car during their weekend strolls through downtown Göttingen (see Fig. 7.6). The plate had been signed by all speakers of this memorial session.

**Fig. 7.6** Photo from the 4th joint ASA and ASJ meeting, Honolulu, 2006, showing the first author in local outfit handing over to Manfred Schroeder a sign that used to indicate the parking spot reserved for goods delivery from the parking lot of the Drittes Physikalische Institut in Göttingen. Photo courtesy of session co-chair Akiro Omoto

## 7.4 Noise Signals with Non-Gaussian Statistics

In the last part of our chapter we want to demonstrate that the use of the very traditional auditory stimulus known as "noise" has also been influenced by advanced signal processing possibilities. Again, the role of mathematics, in this case of amplitude and envelope statistics and of the amount of intrinsic fluctuations and their spectral composition, is evident and has enabled major steps forward in understanding working principles of the human auditory system. Although the meaning of the term noise is wider in daily use, in hearing sciences, it refers to signals that are inherently random. When we consider for example white Gaussian noise, samples taken from its temporal waveform are randomly distributed according to a Gaussian distribution, and samples taken at subsequent moments in time are uncorrelated. The frequency-domain representation of white Gaussian noise shows a complex spectrum where the real and imaginary parts are also normally distributed.

Noise signals are often subjected to some kind of spectral filtering. Although this influences the spectral envelope of the signal, the Gaussian distributions of the time-domain samples and the complex spectral components are not influenced. However, the correlation across samples taken at different moments in time is influenced. This is reflected in the auto-correlation function. For a white noise signal, the autocorrelation function has a peak at lag zero and is zero at all other lags, in line with the idea that samples are mutually uncorrelated. For a filtered noise, however, there will be correlation across samples which is reflected in the autocorrelation function being unequal to zero at non-zero time lags.

Noise signals have a long history in hearing research because they are relatively easy to generate by analog means. Also, the most often used modifications, spectral filtering and temporal windowing, could be realized using basic analog electronics. Noise stimuli have been used extensively to study auditory masking where noise often serves as a masker. For example, in the early experiments related to critical bandwidth by Hawkins and Stevens [23], white, Gaussian-noise maskers were used to measure the frequency dependence of masked thresholds of a tonal signal. The preference for using white noise signals can be attributed to properties such as uniform energy distribution across time and frequency and the fact that the signals can be described by using only a few parameters.

The inherently stochastic nature of noise has strong implications for its masking behavior. This was demonstrated in studies that employed reproducible noise for which the stochastic uncertainties in the noise are effectively removed. Generally, reproducible noise produces lower masked thresholds than running noise (e.g., [39, 77]).

The probability distribution of noise amplitudes and envelopes can also have an influence on the masking properties of a noise signal. As long as stimuli for hearing research were generated by analog means, noise signals typically had a Gaussian statistic. One early exception to this was so-called multiplied noise (also called multiplication, or regular zero-crossing noise), generated by direct multiplication of a sinusoid and a lowpass noise [21]. It was a very convenient way to generate band-pass noises with tunable center frequency, and these noises were therefore quite useful in spectral masking experiments in which noise maskers with variable center frequencies and steep spectral cutoffs were required (e.g., [46, 47]). The fact that the envelope statistics of this stimulus differed significantly from that of Gaussian noise was only recognized after some time, once it was shown that these properties influence the outcome of listening experiments (see, as an example, [76]). One particularly relevant detail, often overlooked, is that the addition of two multiplied noise signals, for which the two sinusoidal center components are uncorrelated, results in a noise with Gaussian statistics.

With the advance of digital computers, additional noise types have been developed, often with the goal to vary the amount and nature of envelope fluctuations in a controlled way. Transposed stimuli were constructed to realize in the envelope of a high-frequency sinusoidal carrier the same waveform that resulted from auditory filtering and half-wave rectification of noise signals with a low center frequency [73]. "Sparse noise" was generated to have a white noise with a controllable, high amount of envelope fluctuations [25]. The opposite in terms of envelope fluctuations was "low-noise noise," a term coined by Pumplin in 1985 [51]. Due to its resemblance with periodic low-peak signals, we will focus in the remaining part of the chapter on this stimulus.

### 7.4.1 *Low-Noise Noise*

#### 7.4.1.1 Definition

In the previous section, we noted that a filtering operation on white Gaussian noise causes the auto-correlation function to change from a delta function at lag zero, to a pattern that reflects predictability of successive time samples of such noise. This predictability is reflected in a smooth development of the envelope of the time-domain noise waveform.[1] The rate of fluctuation in the temporal envelope is proportional to the bandwidth of the filtered noise signal. The spectrum of the envelope has a large DC component and a downward tilting slope that leaves very little spectral power beyond frequencies equal to the bandwidth of the band-pass noise. Interestingly, the *amount* of fluctuation, i.e., the ratio between the AC and the DC part of the envelope, is independent of the bandwidth for Gaussian noise, a property which is reflected in the Rayleigh distribution of the temporal envelope values. Thus, there is an inherently high degree of fluctuation in Gaussian noise.

The inherent fluctuations that are present in Gaussian noise have prompted the development of so-called low-noise noise [51]. This special type of noise has the same spectral envelope as Gaussian noise, but a much lower degree of inherent fluctuations in its temporal envelope, hence the name *low-noise* noise. This stimulus allowed the study of the contribution of envelope fluctuations to auditory masking phenomena by comparing the masking effect of Gaussian and low-noise noise. The first authors to pursue this idea were Hartmann and Pumplin [22].

#### 7.4.1.2 Stimulus Generation

The original means of generating low-noise, such as promoted by [51] was via a special optimization algorithm. First, a band-pass noise was digitally generated in the frequency domain by setting amplitudes in a restricted spectral range to some specific values, e.g., one constant value, and randomizing the phases. Such a noise will approximate all the properties of a band-pass Gaussian noise when the product of duration and bandwidth is sufficiently large. Via a steepest-descent algorithm, the phase spectrum was modified step-by-step in the direction that made the temporal envelope more flat, according to some statistical measure of envelope fluctuation.[2] After a sequence of iterations, a low-noise noise waveform was obtained with a rather flat temporal envelope and the initial amplitude spectrum. Summarizing, the Pumplin's method obtained low-noise noise by modifying the phase spectrum in a special way.

---

[1] There are alternative ways to determine the envelope of a signal which lead to somewhat different envelopes. We will consider here the Hilbert envelope.

[2] In this case, the normalized fourth moment of the waveform.
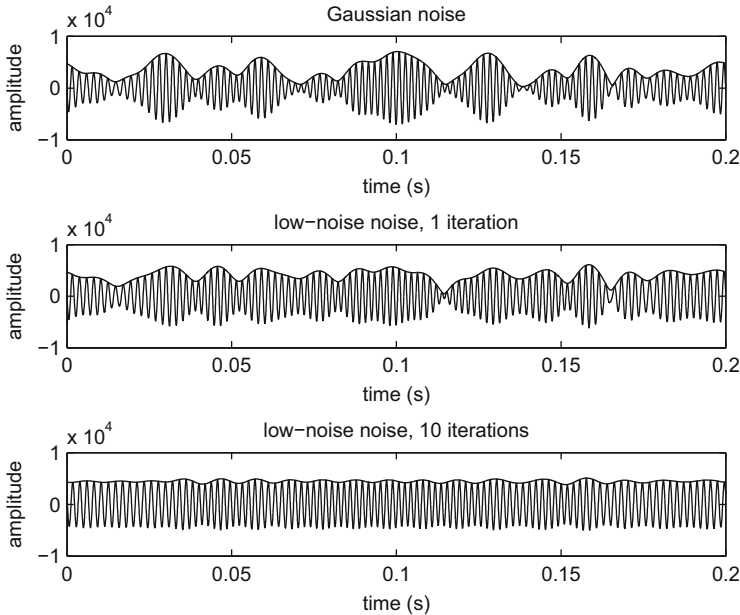
**Fig. 7.7** Illustration of the low-noise noise generation. The *top* panel shows the time-domain Gaussian noise at the start of the iterative process, the *middle* panel the low-noise noise after one iteration, the *lower* panel the low-noise noise after 10 iterations. All waveforms are shown with their respective envelopes

Later on, in a publication dedicated to Manfred Schroeder on the occasion of his 70th birthday, several alternative methods of generating low-noise noise were proposed and evaluated by Kohlrausch et al. [33]. We will here describe the method that led to the lowest degree of fluctuation in the temporal envelope. The method consists of an iterative process that is initiated by generating a time-discrete Gaussian band-pass noise. The iterative process then consists of a sequence of straightforward steps.

First, the Hilbert envelope of the noise is calculated. Secondly, the noise waveform is divided by its Hilbert envelope on a sample-by-sample basis in the time domain. For the rare occasions that the Hilbert envelope is equal to zero, the resulting division is set to zero. In the third step, a band-pass filtering is applied to remove the new spectral components outside of the specified band-pass range that were introduced by the division operation in the previous step. By repeating these steps several times, a much flatter envelope is obtained.

After the first two steps, i.e., after calculating the Hilbert envelope and dividing the noise waveform by this envelope, the resulting waveform will have a flat temporal envelope. The spectrum, however, will also be modified considerably. The division by the Hilbert envelope can be seen as a time-domain multiplication by the reciprocal Hilbert envelope. In the frequency domain, this is equivalent to a convolution of the band-pass noise signal with the spectrum of the reciprocal Hilbert envelope. Due to the large DC component present in the Hilbert envelope,
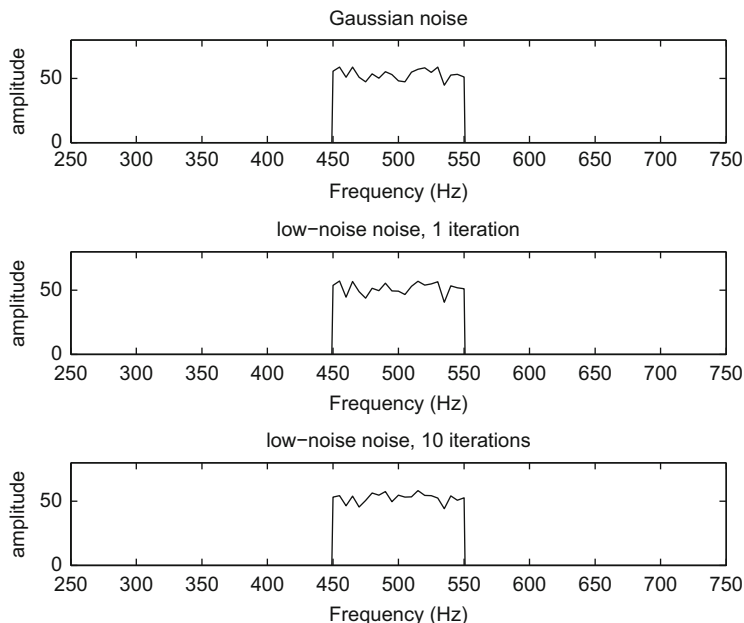
**Fig. 7.8** Illustration of the low-noise noise generation. The *top* panel shows the power spectrum of the Gaussian noise at the start of the iterative process, the *middle* panel the spectrum of the low-noise noise after one iteration, the *lower* panel the spectrum of the low-noise noise after 10 iterations

the reciprocal Hilbert envelope will also have a large DC component. Thus, the convolution in the frequency domain will be dominated by this DC component and as a consequence, the spectrum of the band-pass noise will remain largely intact. However, there will be additional, new spectral components that are outside the band-pass range of the original band-pass noise.

Therefore, in the third step, band-pass filtering is applied to remove the new spectral components outside of the specified band-pass range that were introduced by the division operation in the previous step. Considering the argumentation given above, only a relatively small amount of signal power is removed by this operation. Nevertheless, after filtering, the temporal envelope will not be flat anymore. In Fig. 7.7, the temporal waveforms are shown for the original 100-Hz wide Gaussian band-pass noise centered at 500 Hz, that was input to the iterative process (top panel), and after the first iteration of our algorithm (middle panel). As can be seen, the degree of envelope fluctuation is reduced considerably. By repeating the iterative steps 10 times, a much flatter envelope is obtained (lower panel). Convergence is assumed to be obtained due to the DC component in the Hilbert envelope becoming more dominant over the higher spectral components after each iteration.

In Fig. 7.8, the same signals are shown, only now represented in the frequency domain. As can be seen, the original, band-pass Gaussian noise has a uniform spectral envelope. The spectrum of the low-noise noise signal is, even after

**Table 7.1** Normalized fourth moment of low-noise noise as a function of the number of iterations

| Number of iterations | Normalized fourth moment |
| --- | --- |
| 0 | 3.030 |
| 1 | 1.845 |
| 2 | 1.701 |
| 4 | 1.591 |
| 6 | 1.552 |
| 8 | 1.535 |
| 10 | 1.526 |
| Hartmann and Pumplin | 1.580 |

The bottom row gives the value from [22] for comparison

10 iterations, quite similar to the spectrum of the Gaussian signal. There is, however, a tendency for the spectrum to have a somewhat lower level towards the edges of the band-pass range.

As a measure of envelope fluctuation, Table 7.1 shows the normalized fourth moment for different numbers of iterations of our algorithm. The value obtained by [22] is shown at the bottom of the table. As can be seen, already after 6 iterations, we obtain a lower degree of envelope fluctuation than the method of Hartmann and Pumplin. After 10 iterations, the normalized fourth moment is 1.526, close to the theoretical minimum of 1.5 for a sinusoidal signal.

In summary, the iterative method is able to create a low-noise noise by modifying both the phase and the amplitude spectrum. The specific ordering of spectral components in the passband causes the flat envelope that is seen in the lower panel of Fig. 7.7. Due to the very specific arrangement of phase and amplitude values throughout the noise spectrum, any modification of this spectral ordering will affect the flatness of the temporal envelope. In Fig. 7.9, the low-noise noise signal of Fig. 7.7, which was centered at 500 Hz, and had a bandwidth of 100 Hz, is shown after being filtered with a 78-Hz-wide gamma-tone filter centered at 500 Hz. As can be seen, the degree of envelope fluctuation has increased considerably. Since the gamma-tone filter used here is a reasonable first-order approximation of auditory peripheral filtering, this figure demonstrates that the properties that are present in the external stimulus should not be taken to be representative for the manner in which the stimulus is represented within the auditory system (see also Figs. 6 and 7 in [33]).

### 7.4.2 Role in Hearing Research and Perceptual Insights

As discussed before, Gaussian noise is frequently used as a stimulus in experiments investigating auditory masking. Early experiments by Fletcher [18] used noise signals of various bandwidths to determine detection thresholds of sinusoidal signals centered in the band-pass noise maskers. In these experiments, it was found that only the masker energy that was spectrally close to the sinusoidal target signal contributed to the masking effect of the noise. This led to the concept of the
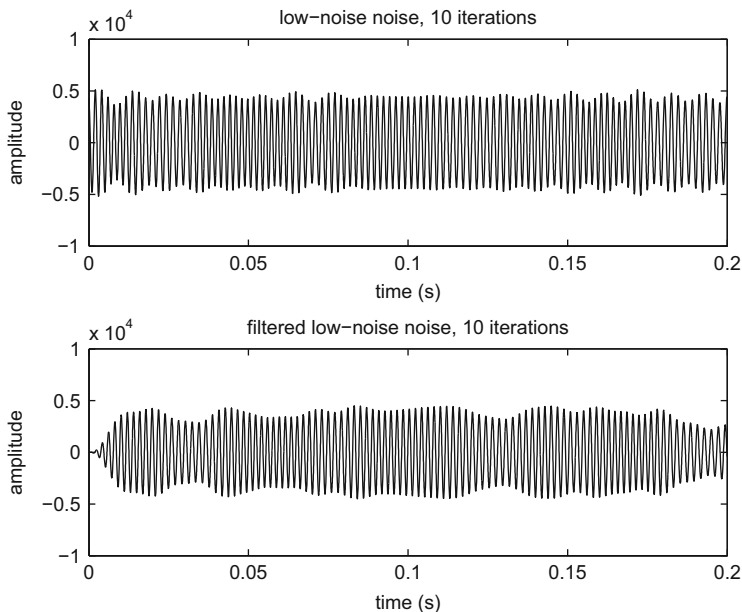
**Fig. 7.9** Illustration of the low-noise noise generation. The *top* panel shows a low-noise noise with a bandwidth of 100 Hz after 10 iterations, the *lower* panel shows the same waveform after peripheral filtering with a gamma-tone filter of 78-Hz width

critical band, which indicates the spectral range that contributes to the masking effect on the sinusoidal signal. The integrated intensity of the masker within this range determines the masked threshold.

This purely intensity-based account of masking does not provide insights into the reasons for observing quite different masked thresholds when narrow-band noises or sinusoidal signals of equal level are used as a masker. When the bandwidth of the noise is smaller than the critical bandwidth, there is no difference in the masker intensity within that critical band for both masker types. Thus if overall masker intensity determines masking, thresholds should be the same. Typically, however, thresholds for tonal maskers are about 20 dB lower than for narrowband Gaussian-noise maskers (e.g., [42]).

One of the factors that is believed to contribute to the different masking strengths of these signals is the difference in the inherent envelope fluctuations. A tonal masker has no inherent envelope fluctuations, and the addition of the target tone will introduce a beating pattern that may be an effective cue for detecting the presence of the target. A noise masker, however, will have a high degree of fluctuation of its own. Addition of the sinusoidal signal does not alter the properties of the envelope fluctuations by a significant degree, and therefore, changes in the temporal envelope pattern may be a less salient cue for a noise masker.

Low-noise noise maskers provide an elegant stimulus to verify that the inherent fluctuations in Gaussian noise are an important factor contributing to its strong

masking effect. Such an experiment had been done by Hartmann and Pumplin [22], but the difference in masked thresholds that they found was only 5 dB. This difference is considerably smaller than the 20-dB difference in masking found for Gaussian-noise maskers and tonal signals. A complicating factor may be that the inherent fluctuations in the low-noise noise that was used by Pumplin and Hartmann were still strong enough to cause a significant masking effect. Furthermore, the bandwidth of their low-noise noise stimulus was 100 Hz around a center frequency of 500 Hz. Although such a bandwidth agrees approximately with the estimates of auditory filter bandwidth at this frequency, peripheral filtering may have caused a significant reduction of the flatness of their low-noise noise stimulus, as we demonstrated in Fig. 7.9, where a low-noise noise signal with their spectral properties was filtered with a 1-ERB wide filter.

A more recent experiment by Kohlrausch et al. [33] used low-noise noise created by the iterative method outlined in the previous section resulting in an even lower degree of inherent fluctuation. In addition, the experiment of Kohlrausch et al. measured masked thresholds as a function of center frequency of the low-noise noise masker while keeping the target tone always spectrally centered within the 100-Hz wide noise masker. The highest center frequency in their experiment was 10 kHz, a frequency where the peripheral filter bandwidth is considerably larger than the masker bandwidth. As a result it can be assumed that peripheral filtering will only have a marginal effect on the temporal envelope flatness of the low-noise noise. Thus in these conditions, the difference in masking thresholds between Gaussian noise and low-noise noise should be about the same size as the difference seen for Gaussian-noise and tonal maskers.

One of the results from [33] is shown in Fig. 7.10. As can be seen, the masked thresholds for Gaussian noise (triangles) are constant for center frequencies of 500 Hz and above. This is in line with the fact that auditory filtering does not reduce masker intensity for a 100-Hz wide masker in this frequency range, and that the degree of inherent envelope fluctuations does not vary as a function of center frequency. For low-noise noise (circles), however, we see a clear dependence of thresholds on the center frequency. Although low-noise noise thresholds were lower than Gaussian-noise thresholds already at a center frequency of 1 kHz, for 10 kHz we see a much larger difference of more than 15 dB, which is much more similar to the difference observed for sinusoidal and Gaussian-noise maskers. The higher thresholds for lower frequencies are well in line with the idea that peripheral filtering affects the temporal envelope flatness of low-noise noise.

A variant of the experiment by Kohlrausch et al. [33] investigated the effect of a frequency offset between masker and sinusoidal target signal with the target always higher in frequency than the masker [75]. The addition of the sinusoidal target to the masker band introduces modulations with a rate that is characterized by the frequency difference between target and masker. When the target is centered within the masker, the newly introduced modulations will have a rate comparable to those already present within the masker alone and will therefore be difficult to detect. When the target is sufficiently remote from the masker band, the modulations that will be introduced due to the addition of the target will be of considerably higher

**Fig. 7.10** Masked thresholds for 100-Hz wide Gaussian-noise (*triangles*) and low-noise noise maskers (*circles*) as a function center frequency. Reprinted from [33] with permission from Hirzel Verlag and the European Acoustics Association

Masked threshold (dB) vs Masker frequency (Hz)

△—△ GN
○—○ LNN
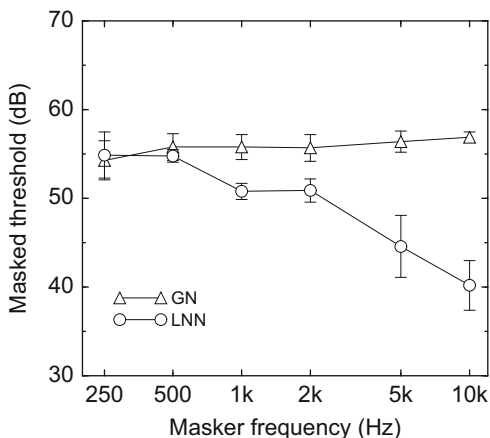
rate than those already present within the masker and may be much easier to detect. Regarding the perceptual processing of stimuli with various modulation rates, there is evidence for frequency selectivity associated with the processing of temporal envelope fluctuations, which led to the modulation filterbank model proposed by [11]. Low-noise noise is a very suitable stimulus to test some of the non-intuitive consequences of this modulation filter bank concept.

In Fig. 7.11, results of the experiment in [75] are summarized. Masked thresholds are shown for Gaussian noise (dashed lines) and low-noise noise (solid lines) maskers centered at 10 kHz for bandwidths of 10 and 100 Hz, indicated by circles and squares, respectively. The abscissa gives the frequency of the sinusoidal target in terms of the frequency offset of the target relative to the masker center frequency. For an abscissa value of 0 Hz, the target is placed at the center frequency of the noise masker. The values at 0 Hz offset represent the most common on-frequency masking conditions. Note that for the two Gaussian-noise maskers (dashed curves), thresholds increase by about 5 dB, when the masker bandwidth is reduced by a factor of 10 from 100 to 10 Hz. This corresponds to a change of 1.5 dB/oct, exactly the value that is predicted by the notion that for this condition the detection process is dominated by an energy cue in combination with increasing variability of this cue for decreasing degrees of freedom in the masker [3, 20, 74].

The corresponding threshold values for low-noise noise are considerably lower, in line with the results from [31], replotted in Fig. 7.10. Again, thresholds for the 10-Hz masker are higher than those for 100 Hz. This can best be explained by the notion of frequency specific modulation perception. The total modulation energy is the same for the two masker bandwidths, but it is concentrated in a much smaller frequency range (basically, within a 10-Hz range of modulation rates) for the 10-Hz masker than for the 100-Hz masker. Thus, the amount of masker modulation within the lowest modulation filter, ranging up to 10 Hz, will be very high for the 10-Hz masker, making it difficult to detect the extra modulations introduced by adding the target (for this argument, see also Figs. 3 and 9 in [31]).

**Fig. 7.11** Masked thresholds for Gaussian-noise (*dashed lines*) and low-noise noise maskers (*solid lines*) as a function of frequency offset between the masker and target. Squares show 100-Hz wide maskers and circles 10-Hz wide maskers both presented at 70 dB SPL
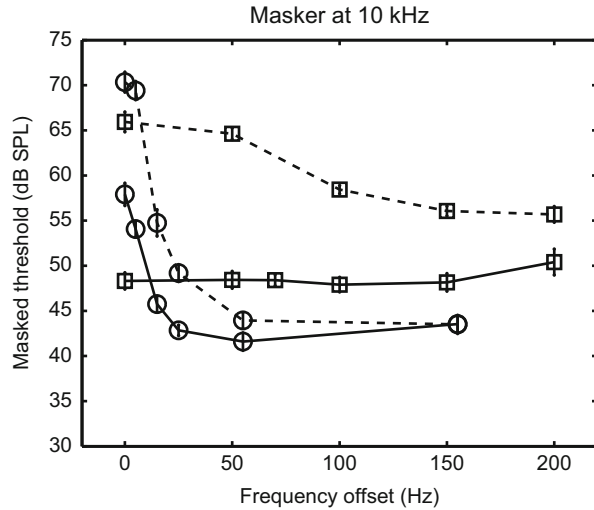


When the target frequency offset is increased, very different patterns are observed for the two different masker bandwidths, but also for the two maskers types that have very different spectra of their intrinsic envelope modulations. For 10-Hz maskers (circles), thresholds start to drop sharply once the offset exceeds 10 Hz. This drop is best explained by the introduction of masker-target modulations with a frequency range exceeding that of the masker modulations. These can be detected the easier, the further they are removed from the masker fluctuations. For the largest offset of 150 Hz, the two maskers lead to identical thresholds, an indication that detection is no longer limited by the intrinsic masker fluctuations ("external noise"), but by the hearing system's limited resolution ("internal noise"). For this condition, the target threshold is about 27 dB below the masker level. This value is in excellent agreement with data for a *sinusoidal* masker which amounted to a threshold about 29 dB below the masker level for a sinusoidal masker at 10 kHz and a sinusoidal target 150 Hz above the masker frequency (see Fig. 8 in [32]).

As can be seen from the squares indicating the 100-Hz-wide maskers, the low-noise noise thresholds (squares with solid lines) are roughly independent of frequency offset. They are about 5 dB higher than the values for the 10-Hz masker, so there is a small, but significant remaining influence of the 100-Hz masker. This can be understood by looking at the envelope spectrum. For a 100-Hz wide masker, the dominant rate of internal envelope fluctuations is around 100 Hz, and this corresponds exactly to the frequency difference between upper edge of the masker and the target for an offset of 150 Hz. Thus, even for frequency offsets of 150 Hz, the relatively small amount of masker envelope fluctuation prevents an optimal detection of the target based on temporal cues. In contrast, the Gaussian-noise thresholds for the same bandwidth (squares with dashed lines) show a clear dependence on the frequency offset, and thresholds are generally higher than the low-noise noise thresholds, in line with the idea that the inherent fluctuations in the

Gaussian masker prohibit the detection of the modulations introduced through the addition of the sinusoidal target signal. At larger frequency offsets, however, the modulations introduced by the target signal become higher in rate, and thresholds decrease systematically. This decrease reflects the decrease of the masker envelope spectrum for a Gaussian noise (cf. Fig. 8 in [15]).

One should note that the frequency offsets discussed here are considerably smaller than the peripheral filter bandwidth at the masker frequency of 10 kHz, which amounts to more than 1,000 Hz. Thus the patterns of thresholds observed in Fig. 7.11 cannot reflect peripheral *spectral* filtering. The dependence of thresholds on target frequency offset that is observed here is a reflection of the processing of temporal envelope fluctuations and not of spectral resolution.

### 7.4.3  Outlook

We have seen that the use of low-noise noise as masker does lead to different thresholds compared to Gaussian noise. This supports the idea that temporal fluctuations in Gaussian noise are a significant factor in auditory masking. Thus, low-noise noise may be an interesting stimulus also in the future to study the contribution of envelope fluctuations to masking.

Low-noise noise may also be of value for acoustical measurement techniques because it is a signal that couples a low crest factor with a continuous spectrum. Whereas in hearing experiments, the bandwidth of low-noise noise is usually limited to, at most, that of one critical bandwidth to prevent peripheral filtering from re-introducing fluctuations in the envelope, for physical measurements this restriction may not exist and wideband low-noise noise may be used to put maximum wideband power into a system that is somehow restricted in its maximum amplitude.

Various studies have applied low-noise noise for a number of different purposes. Dau et al. [15] have used low-noise noise, together with a number of different noise types to study the spectral processing of envelope fluctuations. Due to its flat temporal envelope, low-noise noise has a markedly different envelope spectrum as compared to Gaussian and multiplied noise. The envelope spectral content seems to govern the degree of masking that is observed. Buss et al. [8] have presented low-noise noise in their studies on comodulation masking release, where low-noise-noise provided a masker with little fluctuation but similar bandwidth as a Gaussian-noise masker. Low-noise noise has also been used to measure the minimal level to mask tinnitus [53] because it lacks excessive peaks and troughs. A final application for low-noise noise comes from speech perception research. Healy and Bacon synthesized artificial speech by amplitude modulating bands of low-noise noise and used this type of speech to measure the critical band for speech [24].

## 7.5  Conclusion

The topics in this chapter are naturally biased towards examples, to which we have ourselves contributed, and which were inspired by ideas and publications of Manfred Schroeder. In this way we hope to demonstrate how influential his way of thinking and his early emphasis on using computers in acoustics has been, using the example of psychoacoustics. The great potential of mathematically well-defined acoustic stimulus types lies in the possibility that they can equally well be used in behavioral as well as in physiological experiments, and that they can serve as input to those types of perception models which allow the processing of arbitrary signal waveforms (e.g., [5–7, 11–14]). This close interplay between psychoacoustics, physiology and modeling is one of the central themes of a conference series, the International Symposia on Hearing, which has been mentioned at many places in this chapter. This series was initiated in 1969 by, among others, Manfred R. Schroeder and Jan Schouten, the founding director of the Institute for Perception Research (IPO) in Eindhoven, where the present authors met each other some 20 years ago. The last of the symposia in which Schroeder participated was held 1988 in Paterswolde near Groningen, and Schroeder, with his love for the Dutch people, language, and country was happy to deliver the after dinner address at the conference social outing (see pp. vi–vii in [17]). This symposium never took place in Göttingen, but it can be seen as a late echo of the psychoacoustic research tradition at the DPI that two editions of these symposia, the 12th in 2000 [4] and the 14th in 2006 [38], were co-organized by scientists who received their initial academic training and were shaped in their scientific interests by Manfred Schroeder.

## References

1. Atal, B.S., Schroeder, M.R.: Predictive coding of speech signals and subjective error criteria. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 1978, pp. 573–576 (1978)
2. Blauert, J.: Spatial Hearing. The Psychophysics of Human Sound Localization. The MIT Press, Cambridge, MA (1997)
3. Bos, C.E., de Boer, E.: Masking and discrimination. J. Acoust. Soc. Am. **39**, 708–715 (1966)
4. Breebaart, J., Houtsma, A., Kohlrausch, A., Prijs, V., Schoonhoven, R.: Physiological and Psychophysical Bases of Auditory Function. Shaker Publishers, Maastricht (2001)

5. Breebaart, J., van de Par, S., Kohlrausch, A.: Binaural processing model based on contralateral inhibition. I. Model structure. J. Acoust. Soc. Am. **110**, 1074–1088 (2001)
6. Breebaart, J., van de Par, S., Kohlrausch, A.: Binaural processing model based on contralateral inhibition. II. Predictions as a function of spectral stimulus parameters. J. Acoust. Soc. Am. **110**, 1089–1104 (2001)
7. Breebaart, J., van de Par, S., Kohlrausch, A.: Binaural processing model based on contralateral inhibition. III. Predictions as a function of temporal stimulus parameters. J. Acoust. Soc. Am. **110**, 1105–1117 (2001)
8. Buss, E., Grose, J.H., Hall III, J.W.: Features of across-frequency envelope coherence critical for comodulation masking release. J. Acoust. Soc. Am. **126**, 2455–2466 (2009)
9. Carlyon, R.P., Datta, A.J.: Excitation produced by Schroeder-phase complexes: evidence for fast-acting compression in the auditory system. J. Acoust. Soc. Am. **101**, 3636–3647 (1997)
10. Carlyon, R.P., Datta, A.J.: Masking period patterns of Schroeder-phase complexes: effects of level, number of components, and phase of flanking components. J. Acoust. Soc. Am. **101**, 3648–3657 (1997)
11. Dau, T., Kollmeier, B., Kohlrausch, A.: Modeling auditory processing of amplitude modulation: I. Detection and masking with narrowband carriers. J. Acoust. Soc. Am. **102**, 2892–2905 (1997)
12. Dau, T., Kollmeier, B., Kohlrausch, A.: Modeling auditory processing of amplitude modulation: II. Spectral and temporal integration. J. Acoust. Soc. Am. **102**, 2906–2919 (1997)
13. Dau, T., Püschel, D., Kohlrausch, A.: A quantitative model of the 'effective' signal processing in the auditory system: I. Model structure. J. Acoust. Soc. Am. **99**, 3615–3622 (1996)
14. Dau, T., Püschel, D., Kohlrausch, A.: A quantitative model of the 'effective' signal processing in the auditory system: II. Simulations and measurements. J. Acoust. Soc. Am. **99**, 3623–3631 (1996)
15. Dau, T., Verhey, J., Kohlrausch, A.: Intrinsic envelope fluctuations and modulation-detection thresholds for narrowband noise carriers. J. Acoust. Soc. Am. **106**, 2752–2760 (1999)
16. Duifhuis, H.: Audibility of high harmonics in a periodic pulse. J. Acoust. Soc. Am. **48**, 888–893 (1970)
17. Duifhuis, H., Horst, J.W., Wit, H.P. (eds.): Basic Issues in Hearing. Academic Press, London (1988)
18. Fletcher, H.: Auditory patterns. Rev. Mod. Phys. **12**, 47–65 (1940)
19. Glasberg, B.R., Moore, B.C.J.: Derivation of auditory filter shapes from notched-noise data. Hear. Res. **47**, 103–138 (1990)
20. Green, D.M., Swets, J.A.: Signal Detection Theory and Psychophysics. Wiley, New York, NY (1974). reprinted by Krieger
21. Greenwood, D.D.: Auditory masking and the critical band. J. Acoust. Soc. Am. **33**, 484–502 (1961)
22. Hartmann, W.M., Pumplin, J.: Noise power fluctuation and the masking of sine signals. J. Acoust. Soc. Am. **83**, 2277–2289 (1988)
23. Hawkins, J.E.J., Stevens, S.S.: The masking of pure tones and of speech by white noise. J. Acoust. Soc. Am. **22**, 6–13 (1950)
24. Healy, E.W., Bacon, S.P.: Measuring the critical band for speech. J. Acoust. Soc. Am. **119**, 1083–1091 (2006)
25. Hübner, M., Wiegrebe, L.: The effect of temporal structure on rustling-sound detection in the gleaning bat, Megaderma lyra. J. Comp. Physiol. A **189**, 337–346 (2003)
26. Irino, T., Patterson, R.D.: A time-domain, level-dependent auditory filter: the gammachirp. J. Acoust. Soc. Am. **101**, 412–419 (1997)
27. Johannesma, P.I.M.: The pre-response stimulus ensemble of neurons in the cochlear nucleus. In: Cardozo, B.L., de Boer, E., Plomp, R. (eds.) Proceedings of the 2nd International Symposium on Hearing, pp. 58–69. Institute for Perception Research, Eindhoven (1972)
28. Johnston, J.D.: Transform coding of audio signals using perceptual noise criteria. IEEE J. Selected Areas Commun. **6**, 314–323 (1988)

29. Joris, P.X., van der Heijden, M., Louage, D.H., de Sande, B.V., Kerckhoven, C.V.: Dependence of binaural and cochlear "best delays" on characteristic frequency. In: Pressnitzer, D., de Cheveigné, A., McAdams, S., Collet, L. (eds.) Auditory Signal Processing: Physiology. Psychoacoustics, and Models, pp. 477–483. Springer, New York (2005)

30. Kohlrausch, A.: Masking patterns of harmonic complex tone maskers and the role of the inner ear transfer function. In: Duifhuis, H., Horst, J., Wit, H. (eds.) Basic Issues in Hearing, pp. 339–350. Academic Press, Harcourt Brace Jovanovich. Publ, London (1988)

31. Kohlrausch, A., Fassel, R.: Binaural masking level differences in nonsimultaneous masking. In: Gilkey, R.H., Anderson, T. (eds.) Binaural and Spatial Hearing in Real and Virtual Environments, Chapter 9, pp. 169–190. Lawrence Erlbaum Ass, London (1997)

32. Kohlrausch, A., Fassel, R., Dau, T.: The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers. J. Acoust. Soc. Am. **108**, 723–734 (2000)

33. Kohlrausch, A., Fassel, R., van der Heijden, M., Kortekaas, R., van de Par, S., Oxenham, A.J., Püschel, D.: Detection of tones in low-noise noise: further evidence for the role of envelope fluctuations. Acustica United Acta Acustica **83**, 659–669 (1997)

34. Kohlrausch, A., Houtsma, A.J.M.: Edge pitch of harmonic complex tones. IPO Annual Progress Report **26**, 39–49 (1991)

35. Kohlrausch, A., Houtsma, A.J.M.: Pitch related to spectral edges of broadband signals. Phil. Trans. R. Soc. Lond. B **336**, 375–382 (1992)

36. Kohlrausch, A., Sander, A.: Phase effects in masking related to dispersion in the inner ear. II. Masking period patterns of short targets. J. Acoust. Soc. Am. **97**, 1817–1829 (1995)

37. Kohlrausch, A., van de Par, S.: On the use of specific signal types in hearing research. In: Kurtz, T., Parlitz, U., Kaatze, U. (eds.) Oscillations, Waves, and Interactions. Sixty Years Drittes Physikalisches Institut. A Festschrift, pp. 37–71. Universitätsverlag Göttingen, Göttingen (2007)

38. Kollmeier, A., Klump, G., Hohmann, V., Langemann, U., Mauermann, M., Verhey, J. (eds.): Hearing—From Sensory Processing to Perception. Proceedings of the 14th International Symposium on Hearing. Springer, Berlin (2007)

39. Langhans, A., Kohlrausch, A.: Differences in auditory performance between monaural and diotic conditions. I: Masked thresholds in frozen noise. J. Acoust. Soc. Am. **91**, 3456–3470 (1992)

40. Lentz, J.J., Leek, M.R.: Psychophysical estimates of cochlear phase response: masking by harmonic complexes. J. Assn. Res. Otolaryngol. **2**, 408–422 (2001)

41. Mehrgardt, S., Schroeder, M.R.: Monaural phase effects in masking with multicomponent signals. In: Klinke, R., Hartmann, R. (eds.) Hearing—Physiological Bases and Psychophysics, pp. 289–295. Springer, New York (1983)

42. Moore, C., Alcántara, J.I., Dau, T.: Masking patterns for sinusoidal and narrow-band noise maskers. J. Acoust. Soc. Am. **104**, 1023–1038 (1998)

43. Münkner, S., Kohlrausch, A., Püschel, D.: Influence of fine structure and envelope variability on gap-duration discrimination thresholds. J. Acoust. Soc. Am. **99**, 3126–3137 (1996)

44. Oxenham, A.J., Dau, T.: Reconciling frequency selectivity and phase effects in masking. J. Acoust. Soc. Am. **110**, 1525–1538 (2001)

45. Oxenham, J., Dau, T.: Towards a measure of auditory-filter phase response. J. Acoust. Soc. Am. **110**, 3169–3178 (2001)

46. Patterson, R.D.: Auditory filter shape. J. Acoust. Soc. Am. **55**, 802–809 (1974)

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

48. Patterson, R.D., Allerhand, M.H., Giguère, C.: Time-domain modelling of peripheral auditory processing: a modular architecture and a software platform. J. Acoust. Soc. Am. **98**, 1890–1894 (1995)

49. Patterson, R.D., Nimmo-Smith, I., Weber, D.L., Milroy, R.: The deterioration of hearing with age: frequency selectivity, the critical ratio, the audiogram, and speech threshold. J. Acoust. Soc. Am. **72**, 1788–1803 (1982)

50. Pierce, J.R.: Some work on hearing. Am. Scientist **48**, 40–45 (1960)
51. Pumplin, J.: Low-noise noise. J. Acoust. Soc. Am. **78**, 100–104 (1985)
52. Recio, A., Rhode, W.S.: Basilar membrane responses to broadband stimuli. J. Acoust. Soc. Am. **108**, 2281–2298 (2000)
53. Schneider, P., Andermann, M., Wengenroth, M., Goebel, R., Flor, H., Rupp, A., Diesch, E.: Reduced volume of Heschl's gyrus in tinnitus. Neuroimage **45**, 927–927 (2009)
54. Schroeder, M.R.: New results concerning monaural phase sensitivity. J. Acoust. Soc. Am. **31**, 1579 (1959)
55. Schroeder, M.R.: Vocoders: analysis and synthesis speech. Proc. IEEE **54**, 720–734 (1966)
56. Schroeder, M.R.: Relation between critical bands in hearing and the phase characteristics of the cubic difference tones. J. Acoust. Soc. Am. **46**, 1488–1492 (1969)
57. Schroeder, M.R.: Synthesis of low-peak-factor signals and binary sequences with low autocorrelation. IEEE Transact. Inf. Theor. **16**, 85–89 (1970)
58. Schroeder, M.R.: An integrable model for the basilar membrane. J. Acoust. Soc. Am. **53**, 429–434 (1973)
59. Schroeder, M.R.: Models of hearing. Proc. IEEE **63**, 1332–1350 (1975)
60. Schroeder, M.R.: New viewpoints in binaural interaction. In: Evans, E.F., Wilson, J.P. (eds.) Psychophysics and Physiology of Hearing, pp. 455–456. Academic Press, New York (1977)
61. Schroeder, M.R., Atal, B.S.: Computer simulation of sound transmission in rooms. IEEE Int. Convention Record **11**, 150–155 (1963)
62. Schroeder, M.R., Atal, B.S., Hall, J.: Optimizing digital speech coders by exploiting masking properties of the human ear. J. Acoust. Soc. Am. **66**, 1647–1652 (1979)
63. Schroeder, M.R., Atal, B.S., Sessler, G.M., West, J.E.: Acoustical measurements in Philharmonic Hall (New York). J. Acoust. Soc. Am. **40**, 434–440 (1966)
64. Schroeder, M.R., Gottlob, D., Siebrasse, K.F.: Comparative study of European concert halls. J. Acoust. Soc. Am. **56**, 1195–1201 (1974)
65. Schroeder, M.R., Hall, J.L.: Model for mechanical to neural transduction in the auditory receptor. J. Acoust. Soc. Am. **55**, 1055–1060 (1974)
66. Schroeder, M.R., Strube, H.W.: Flat-spectrum speech. J. Acoust. Soc. Am. **79**, 1580–1583 (1986)
67. Shamma, S.A., Shen, N., Gopalaswamy, P.: Stereausis: Binaural processing without neural delays. J. Acoust. Soc. Am. **86**, 989–1006 (1989)
68. Smith. Der Einfluß des Vorzeichens des Momentanfrequenzverlaufes von Maskierern auf die Detektion von reinen Tönen (The influence of the sign of the change in the instantaneous frequency of a masker on the detectability of pure tones). Master's Thesis, Georg-August-Universität Göttingen (1984)
69. Smith, K., Sieben, U., Kohlrausch, A., Schroeder, M.R.: Phase effects in masking related to dispersion in the inner ear. J. Acoust. Soc. Am. **80**, 1631–1637 (1986)
70. Strube, H.W.: A computationally efficient basilar-membrane model. Acustica **58**, 207–214 (1985)
71. Summers, V.: Effects of hearing impairment and presentation level on masking period patterns for Schroeder-phase harmonic complexes. J. Acoust. Soc. Am. **108**, 2307–2317 (2000)
72. Summers, V., Leek, M.R.: Masking of tones and speech by Schroeder-phase harmonic complexes in normally hearing and hearing-impaired listeners. Hear. Res. **118**, 139–150 (1998)
73. van de Par, S., Kohlrausch, A.: A new approach to comparing binaural masking level differences at low and high frequencies. J. Acoust. Soc. Am. **101**, 1671–1680 (1997)
74. van de Par, S., Kohlrausch, A.: Diotic and dichotic detection using multiplied-noise maskers. J. Acoust. Soc. Am. **103**, 2100–2110 (1998)
75. van de Par, S., Kohlrausch, A.: The role of intrinsic masker fluctuations on the spectral spread of masking. In *Forum Acusticum 2005, Budapest*, pp. 1635–1640, European Acoustics Association (2005)

76. van der Heijden, M., Kohlrausch, A.: The role of envelope fluctuations in spectral masking. J. Acoust. Soc. Am. **97**, 1800–1807 (1995)
77. von Klitzing, R., Kohlrausch, A.: Effect of masker level on overshoot in running- and frozen-noise maskers. J. Acoust. Soc. Am. **95**, 2182–2201 (1994)

<div align="center">***</div>

## Biography



**Armin Kohlrausch** studied physics at the University of Göttingen (Germany) and specialized in acoustics. He received his master degree in 1980 and his Ph.D. in 1984, both on perceptual aspects of sound. From 1985 until 1990 he worked at the Drittes Physikalisches Institut at the University of Göttingen, being responsible for research and teaching in the fields psychoacoustics and room acoustics. In 1991 he joined the Philips Research Laboratories in Eindhoven and worked in the speech and hearing group of the Institute for Perception Research (IPO), a joint venture between Philips and the Eindhoven University of Technology (TU/e). Since 1998, he combines his work at Philips Research Laboratories with a Professor position for auditory and multisensory perception at the TU/e. In 2004 he was appointed research fellow of Philips Research. He is a member of a great number of scientific societies, both in Europe and in the USA. Since 1998 he is a Fellow of the Acoustical Society of America and has served as Associate Editor for the Journal of the Acoustical Society of America, for the Journal of the Association for Research in Otolaryngology and for Acta Acustica United with Acustica. His main scientific interest is the experimental study and modeling of auditory and multisensory perception in humans and the transfer of this knowledge to industrial applications.



**Steven van de Par** studied physics at the Eindhoven University of Technology, Eindhoven, The Netherlands, and received the Ph.D. degree in 1998 from the Eindhoven University of Technology, on a topic related to binaural hearing. As a Postdoctoral Researcher at the Eindhoven University of Technology, he studied auditory–visual interaction and was a Guest Researcher at the University of Connecticut Health Center. In early 2000, he joined Philips Research, Eindhoven, to do applied research in auditory and multisensory perception, low-bit-rate audio coding and music information retrieval. Since April 2010 he holds a professor position in acoustics at the University of Oldenburg, Germany, with a research

focus on the fundamentals of auditory perception and its application to virtual acoustics, vehicle acoustics, and digital signal processing. He has published various papers on binaural auditory perception, auditory–visual synchrony perception, audio coding, and computational auditory scene analysis.