Signal detection in fluctuating background noise is a common problem in diverse fields of research and technology. It has been shown in hearing research that the detection of signals in noise that is correlated in amplitude across the frequency spectrum (comodulated) can be improved compared to uncorrelated background noise. We show that the mechanism leading to this effect is a general phenomenon which may be utilized in other areas where signal detection in comodulated noise needs to be done with a limited frequency resolution. Our model is based on neurophysiological experiments. The proposed signal detection scheme evaluates a fluctuating envelope, the statistics of which depend on the correlation structure across the spectrum of the noise. In our model, signal detection does not require a sophisticated neuronal network but can be accomplished through the encoding of the compressed stimulus envelope in the firing rate of neurons in the auditory system.

Keywords: Amplitude fluctuations; comodulation detection difference; comodulation masking release; auditory system.

1. Introduction

Deciding if a signal is present in background noise is the central problem of signal detection. The statistical properties of noise masking a signal depend on the origin of fluctuations: the standard property of Gaussian white noise is a consequence of the central limit theorem when assuming the fluctuations to arise from
a superposition of many independent individual impacts, as in Brownian motion, whereas exponentially correlated red noise results from the Ornstein-Uhlenbeck process [1]. A not so frequently investigated but nonetheless important situation is the case of *comodulated noise*. It occurs when amplitudes at different regions of the frequency spectrum fluctuate in a coherent, or temporally correlated, fashion. This *comodulation* can be generated when sound pressure waves propagate through turbulent air [2] or as a result of the biomechanics of sound production in humans and other animals [3]. Random amplitude fluctuations that may be correlated across the frequency spectrum occur also in other natural phenomena, for example when (star-)light or radio waves are influenced by turbulences in the atmosphere [4], and in engineering and technology applications, such as radar [5] or sonar [6]. Hence, the detection of signals in comodulated noise is a problem of general interest across disparate disciplines.

The auditory system continually performs signal detection tasks. One familiar example is speech perception in noisy social environments, such as a cocktail party [7]. Background noise that is hampering the detection of a signal can be called masking noise. We believe a better understanding of signal detection can be achieved by modeling these processes in the vertebrate auditory system, which appears to have evolved to exploit comodulated noise for lowering detection thresholds [8]. Here, we propose such a model based on neurophysiological recordings from the auditory forebrain of the European starling (*Sturnus vulgaris*).

Fig. 1. Spectrogram of a typical stimulus for the co-uncorrelated (CU) condition similar to the ones presented to starlings in the experiments. The signal band centered around 2 kHz has a different envelope (i.e. amplitude modulations) than the six masker bands which share a common envelope.

In both humans and starlings, thresholds for detecting acoustic signals in noise can be lower in comodulated noise compared to unmodulated noise or incoherently modulated noise with independent amplitude fluctuations across the spectrum [9]. One example of this general phenomenon is the comodulation detection difference (CDD) [10]. Experimental tests of the CDD effect in humans [10,11] and starlings [12] have shown that the detection of an amplitude modulated narrowband noise signal is improved when the amplitude envelopes of several spectrally separated narrowband noise maskers are correlated with each other (i.e., comodulated) and independent of that of the signal band (the co-uncorrelated [CU] condition; see
Fig. 1). This improvement in signal detection (i.e., lower detection thresholds) is relative to conditions in which the signal band and each masker band have either a common envelope (the all-correlated [AC] condition) or independent envelopes (the all-uncorrelated [AU] condition). Thresholds in the CU condition can be up to 10 dB lower than those in the AC and AU conditions, which are generally similar [10–12].

The inner ear functions as a bank of bandpass filters, or channels, in the frequency domain. In the present work, we describe a simple model that uses the output of only one such channel to account for CDD, which is consistent with current psychophysical findings [11]. Our model is based on neurophysiological data recorded in recent CDD experiments with starlings [12] and focuses on a statistical evaluation of the fluctuating envelope of a stimulus comprised of a signal band and a number of masker bands (see Fig. 1). The model developed below can predict a number of qualitative and quantitative results from experimental studies of the CDD effect. Moreover, it is based on simple, well-known peripheral processing stages in the auditory system that could be implemented easily into signal detection schemes in other fields. The main stages are frequency filtering, calculation of the envelope, and compression (i.e. a nonlinear transform of the envelope).

2. Experimental Background

Electrophysiological recordings were made from six awake and freely moving starlings (three adult females, three adult males) that did not have to perform any specific tasks. Descriptions of the experimental setup and procedure can be found in [12] and will be detailed in a future paper. The mean firing rates of small populations of neurons (about 5 units) were obtained from extracellular recordings of field L2 (an equivalent of the mammalian primary auditory cortex) during repeated presentations of stimuli similar to the one shown in Fig. 1. The central frequency of the signal band was set to the frequency at which the neurons were maximally sensitive. The central frequencies of adjacent noise bands (signal and masker bands) were separated by 300 Hz, and each noise band had a bandwidth of 100 Hz. The level of the masker bands was held constant across all conditions. Overall masker levels of 42.8 dB SPL (sound pressure level) and 77.8 dB SPL were tested. For each of the possible stimulus configurations, 30 random stimuli were generated and played back with a 700 ms silent period between two stimuli. The first 20 artifact free recordings were used to determine the average firing rate of the neurons during the time of signal presentation, taking into account the delay between stimulus playback and onset of neuronal activity (on the order of 15 ms). As the level of the signal band was increased in 5 dB steps, the time- and ensemble-averaged firing rates showed different patterns in the three correlation conditions (Fig. 2A). Two important trends in these data are worth noting. First, neural responses in the AU-condition exhibited higher mean firing rates at low signal levels compared to the other two conditions. Second, the firing rates in the AC-condition exhibited a clearly nonmonotonic pattern. These general findings were found for all birds, two different signal durations (either 60 ms or 400 ms) and two different signal offsets (either 0 ms (i.e. maskers and signal starting at the same time) or 100 ms). Signal detection thresholds for the physiological data were determined by calculating the discriminability index $d'$, which is defined as the difference of the mean spike rates
when the signal is present and when it is absent divided by their common standard deviation, assuming both standard deviations are equal [13]. Our experimental results showed that the firing rate standard deviations do not depend strongly on signal level. For the reasonable empirical criterion $d' = 1.8$ [14] the signal is said to have been detected (Fig. 2B). The rank order of detection thresholds (CU < AC ≈ AU) is consistent with perceptual studies of humans [10,11] and starlings [12].

3. Model and Results

The firing rate of auditory neurons can encode the envelope of sound stimuli [15]. Therefore, we assume that the time- and ensemble-averaged firing rates observed in our physiological recordings are a measure of the mean stimulus envelope during signal presentation. Our model translates the mean envelope of a stimulus into a neuronal firing rate. The stimuli that serve as the input to our model are comprised of a number of narrowband noises like the ones shown in Fig. 1. The first step of the model consists of spectrally filtering the stimulus by decreasing the amplitude of the

---

**Fig. 2.** A: Spike discharge rates from a typical recording site averaged across time and trials and normalized to the average rate if maskers alone are present in the AC condition. B: $d'$-analysis of spike discharge rates. C: Expectation values of the time- and ensemble-averaged compressed envelope amplitudes relative to their value if maskers alone are present in the AC condition. A saturating function was applied to account for saturation of the neural firing rate. D: $d'$-analysis of compressed saturating envelopes. For the model, a constant standard deviation $\sigma = 0.18$ in normalized units was assumed. The solid horizontal line in graphs B and D indicates the detection criterion of $d' = 1.8$. (Model parameters used for all figures if not stated otherwise: $\alpha = 0.35$, $a_1 = a_{-1} = 0.8$, $a_2 = a_{-2} = 0.5$, $a_3 = a_{-3} = 0.25$. )
masker bands with increasing distance from the signal band. This filtering mimics the frequency selectivity of the inner ear. The filtered stimulus can be understood as the real part of the analytical signal [16]

\[
s(t) = \sum_{k=-M}^{M} a_k \sum_{n=-N}^{N} e^{i(\omega_0 + k\Delta\omega + n\Delta\omega) t + \phi_{k,n}}.
\]

Here, \(2M\) is the number of masker bands, \(2N + 1\) is the number of sinusoids used to generate the noise bands, \(\omega_0\) is the central frequency of the signal, and \(\Delta\omega\) and \(\Delta\omega\) define the spacing between noise bands and component sinusoids, respectively (Fig. 3). The real factors \(a_k\) denote the filtered amplitudes of the different noise bands and the signal amplitude is determined by \(a_0\). Frequency selectivity is modeled by decreasing \(a_k\) for increasing \(|k|\). We use a symmetric filter \((a_{-k} = a_k)\) with a realistic bandwidth for the stargle auditory system [17]. The phases \(\phi_{k,n}\) are the distinguishing element for the different stimulus conditions: if all \(\phi_{k,n}\) are randomly chosen from \([0, 2\pi]\), then an AU-stimulus is generated; if \(\phi_{k,n}\) are random for all \(n\) but equal for all \(k\), then an AC stimulus is generated; CU stimuli are similar to AC stimuli except for the fact that only for \(k, k' = 0\) the phases \(\phi_{k,n}\) and \(\phi_{k',n'}\) are the same.

The envelope of the filtered stimulus can be easily calculated as the absolute value of the analytical signal \(s(t)\). For assessing the envelope statistics, first of all the ensemble average \(\langle |s(t)|^2 \rangle_\phi\) can be determined. Using trigonometric addition formulas, one can transform Eq. 1 to

\[
|s(t)|^2 = \sum_{k,k'=\pm M} a_ka_{k'} \sum_{n,n'=N} \cos((k-k')\Delta\omega + (n-n')\Delta\omega) t + \phi_{k,n} - \phi_{k',n'}).
\]

Making use of further addition theorems and keeping in mind that

\[
\langle \cos(\phi_{k,n}) \sin(\phi_{k',n'}) \rangle_\phi = \langle \sin(\phi_{k,n}) \cos(\phi_{k',n'}) \rangle_\phi = 0,
\]

\[
\langle \cos(\phi_{k,n}) \cos(\phi_{k',n'}) \rangle_\phi = 0 \text{ if } n \neq n',
\]

\[
\langle \cos(\phi_{k,n}) \cos(\phi_{k,n}) \rangle_\phi = \frac{1}{2},
\]
one arrives at

$$\langle |s(t)|^2 \rangle_\phi = (2N + 1) \sum_{k,k'=-M}^{M} a_k a_{k'} \delta(\phi_{k,n}, \phi_{k',n}) \cos((k - k')\Delta \Omega t). \quad (6)$$

The \(\delta\)-terms are one if \(\phi_{k,n} = \phi_{k',n}\) and zero if \(\phi_{k,n} \neq \phi_{k',n}\). Note that by taking the mean over all possible phases \(\phi\), the dependence on \(\Delta \Omega\) disappears while \(|s(t)|^2\) is already independent of the central frequency \(\omega_0\). For large \(N\), the distribution of squared envelope amplitudes at a certain time \(t\) can be approximated as an exponential distribution (see e.g. [18]). Assuming symmetrical filtering \((a_k = a_{-k})\), the mean of this distribution for the three correlation conditions at time \(t\) is given by

$$\mu^2_{\Delta u} = \langle |s(t)|^2 \rangle_{\phi,\Delta u} = (2N + 1) \left[ a_0^2 + 2 \sum_{k=1}^{M} a_k^2 \right], \quad (7)$$

$$\mu^2_{\Delta c}(t) = \langle |s(t)|^2 \rangle_{\phi,\Delta c} = (2N + 1) \left[ a_0 + 2 \sum_{k=1}^{M} a_k \cos(k\Delta \Omega t) \right]^2, \quad (8)$$

$$\mu^2_{cu}(t) = \langle |s(t)|^2 \rangle_{\phi,cu} = (2N + 1) \left[ a_0^2 + 4 \left( \sum_{k=1}^{M} a_k \cos(k\Delta \Omega t) \right)^2 \right]. \quad (9)$$

Here, the correlations of the phases were used to substitute the \(\delta\)-terms. If the distribution of squared envelope values is an exponential distribution with parameter \(\mu^2(t)\), then the distribution of envelope amplitudes \(Y\) at time \(t\) is given by a Rayleigh distribution with mean \(\sqrt{2\mu(t)}\):

$$p_Y(y; t) = \frac{2y}{\mu^2(t)} \exp\left(-\frac{y^2}{\mu^2(t)}\right) \quad (10)$$

Making use of separate time scales (fast fluctuations with multiples of \(\Delta \Omega\) and slow fluctuations with multiples of \(\Delta \omega\)) and a stimulus duration that is much larger than \(2\pi/\Delta \Omega\), the final envelope distribution and its moments can be computed by eliminating time. This is done by integrating the time dependent Rayleigh distribution over one period of its parameter \(\mu^2(t)\). With \(Y\) being the random variable for the envelope amplitude, one gets

$$E(Y^\alpha) = \int_0^\infty y^\alpha \langle p_Y(y; t) \rangle_T \, dy$$

$$= \Gamma\left(\frac{\alpha + 2}{2}\right) \frac{1}{T} \int_0^T |\mu(t)|^\alpha \, dt \quad (11)$$

as the expectation value of \(Y^\alpha\), where \(T = \frac{2\pi}{\Delta \Omega}\) denotes the duration of one period of \(\mu^2(t)\). This expression holds for any \(\alpha > 0\) and can be used to compute all moments of the distribution of \(Y^\alpha\).

A compressive nonlinearity that is already present in the inner ear [19] is implemented in our model by taking the envelope of the filtered stimulus to the power of
\[ E(Y^a) \]

\[ \text{Std}(Y^a) \]

Fig. 4. Mean (A) and standard deviation (B) of envelope distributions for increasing signal level and constant masker level as predicted from Eq. 11 (lines) and as found in sample stimuli (markers) for \( M = 3, N = 50, \Delta \Omega/2\pi = 300 \text{ Hz}, \omega_0/2\pi = 2000 \text{ Hz}, \) and \( \Delta \omega/2\pi = 1 \text{ Hz}. \) A compressive nonlinearity is accounted for by the exponent \( \alpha = 0.35. \) Variability in the data for sample stimuli results from a finite stimulus duration (400 ms).

\[ \alpha \] with \( \alpha < 1 \) (see Eq. 11). We use a compression of \( \alpha = 0.35, \) which is consistent with our own neural data and recordings from the avian auditory nerve [19]. Mean and standard deviation (SD) of the envelope distributions for this compression are shown in Fig. 4.

In our model, the translation of compressed envelope values to neuronal firing rates is assumed to be linear up to a certain value. Then saturation is taken into account by an empirical function. The average firing rate can be computed as proportional to the time and ensemble expectation value of the compressed saturating envelope of the filtered stimulus. The output of this model for realistic parameters is shown in Fig. 2C. Saturation is needed for comparison with the experimentally obtained firing rates and can be omitted without altering the general effect of different thresholds for the three correlation conditions. Eq. 11 can be used in the case of no saturation.

To estimate the magnitude of the CDD effect predicted by the model, we implemented the same signal detection scheme that was used to determine physiological thresholds. We determined \( d' \) by assuming a firing rate standard deviation \( \sigma \) independent of signal level and stimulus condition. This is consistent with our neural data. Due to this independence, our detection scheme reduces to a comparison of firing rates in the conditions with and without signal: a certain increase in firing rate indicates signal detection. As shown in Fig. 2D, the model predicts thresholds for the CU, AC, and AU conditions that are similar to those depicted for the example recording site (Fig. 2B). The rank order of thresholds agrees with previous CDD experiments in humans [10, 11] and starlings [12]. The amount of CDD as determined from our experiments and our model ought to be less than found in psychophysics: The firing rate standard deviation of small neural populations is larger than that of large populations of neurons. A smaller standard deviation for the calculation of \( d' \) results in a vertical expansion of the curves in Fig. 2D and therefore yields an increase in CDD if the detection criterion of \( d' = 1.8 \) is kept constant.
The threshold differences between the three correlation conditions can be understood by noting two important differences in the mean compressed envelopes (Figs. 2 and 4A). First, there is a prominent "dip" in the AC condition (present for $0<\alpha<2$). This dip results from the beat phenomenon and occurs when the overall signal and masker levels are approximately equal. Here, beating due to correlated phases has the biggest effect. This dip causes the corresponding AC discriminability index ($d'$) to exceed the detection criterion at a higher signal level compared to the CU condition (Fig. 2D). Second, the mean envelope of a stimulus comprised only of the maskers ($a_0 = 0$) is higher in the AU condition than in the other two conditions (Fig. 2C). This result is also rooted in the beat phenomenon. In the AU condition, the incoherently modulated masker bands do not interfere strongly with each other, while in the AC and CU conditions interference due to common phases leads to a relative reduction of the average compressed envelope. This explains why $d'$, which is computed based on the difference between the signal plus masker and masker alone envelopes, crosses the threshold criterion at a higher signal level in the AU condition compared to the CU condition.

4. Conclusions
Despite its widespread presence in nature and in technology applications, comodulated noise has not been widely discussed outside the hearing research literature. Models of signal detection in comodulated noise inspired by the vertebrate auditory system have potentially far reaching implications for other fields in which signal detection in comodulated noise is a general problem. As an example, we proposed a model for a well-known hearing phenomenon (CDD) in which signal detection is enhanced by comodulated noise. Using single channel processing of the mean stimulus envelope, our model successfully reproduces neural responses and threshold differences demonstrated in experiments of the CDD effect. The reason for different envelope statistics in the CU, AC, and AU conditions is the beat phenomenon. The main effect of differing detection thresholds remains even if frequency filtering and saturation are omitted; the essential steps are envelope extraction and compression. This insight may be transferable to other fields, where the frequency resolution of the detection device is insufficient for the signal to be detected by standard filtering. In a situation like this, the compressed envelope might be a suitable quantity for analysis. We see potential applications in any field where comodulated noise may occur (e. g. radar, sonar, ultrasound, wireless communications, or seismic measurements).

Acknowledgements
We thank Jesko Verhey for stimulating discussions. This study was supported by DFG GRK 591, DFG FG 306, and NSF INT-0107304.

References