A quantitative model of the “effective” signal processing in the auditory system. I. Model structure

Torsten Dau\(^a\) and Dirk Püschel\(^b\)

\(^a\)Drittes Physikalisches Institut, Universität Göttingen, Bürgerstr. 42-44, 37073 Göttingen, Germany

\(^b\)Institute for Perception Research (IPO), P.O. Box 513, 5600 MB Eindhoven, The Netherlands

(Received 20 June 1994; revised 24 January 1996; accepted 18 February 1996)

This paper describes a quantitative model for signal processing in the auditory system. The model combines a series of preprocessing stages with an optimal detector as the decision device. The present paper gives a description of the various preprocessing stages and of the implementation of the optimal detector. The output of the preprocessing stages is a time-varying activity pattern to which “internal noise” is added. In the decision process, a stored temporal representation of the signal to be detected (template) is compared with the actual activity pattern. The comparison amounts to calculating the correlation between the two temporal patterns and is comparable to a “matched filtering” process. The detector itself derives the template at the beginning of each simulated threshold measurement from a suprathreshold value of the stimulus. The model allows one to estimate thresholds with the same signals and psychophysical procedures as those used in actual experiments. In the accompanying paper [Dau et al., J. Acoust. Soc. Am. 99, 3623–3631 (1996)] data obtained for human observers are compared with the optimal-detector model for various masking conditions. © 1996 Acoustical Society of America.

PACS numbers: 43.66.Ba, 43.66.Dc

INTRODUCTION

This is the first of two papers in which we describe a quantitative model for signal processing in the auditory system. This model combines several stages which simulate aspects of transformation in the auditory periphery with an optimal detector as the decision device.

The model described in this paper is applicable to masking conditions with stochastic as well as with deterministic masks, independent of the temporal relation between masker and signal. In comparison to other models that allow the computation of masked thresholds, the present model differs in two relevant aspects. As part of the peripheral processing part, it contains a stage that simulates aspects of temporal adaptation. This part is responsible for the ability of the model to predict the decay of thresholds in forward-masking conditions (Püschel, 1988). The decision stage is realized as an optimal detection process which is based on the signal detection theory (Green and Swets, 1966). This choice for the decision stage allows the validity of the preprocessing stages to be assessed. If, under a specific condition, the model is systematically less sensitive than real observers, we can conclude that too much information is lost in the preprocessing parts of the model. Thus discrepancies of this kind are clear hints as to where the model has to be changed, and examples for this will be presented in the accompanying paper.

One main property of the model is that it deals with actual time signals and that threshold values are derived using an adaptive tracking procedure in a simulated three-interval forced-choice procedure. Therefore, the same stimuli as in psychoacoustic experiments can be used in the simulations, and the results of the optimal processor model can be directly compared with the thresholds obtained by human observers.

In this first paper we will describe the specific properties of the various model stages. In this description, we concentrate on aspects in which the present model differs from other models. A comparison of model predictions with the performance of human observers is provided in the accompanying paper.

I. DESCRIPTION OF THE MODEL

A. General aspects

The model described in this paper consists of two major parts, which will be described in detail in the two following Secs. 1B and 1C. In the first part of the model (which includes the blocks labeled “preprocessing,” “adaption,” and “internal noise” in Fig. 1), several stages of the auditory periphery are simulated. Through the transformations occurring in these stages, certain details in the incoming sounds are lost, like, e.g., the temporal fine structure in high-frequency auditory channels. The internal representation of auditory stimuli at the output of this part is then analyzed by an optimal detection device reflecting the second part of the model. The detector is designed to make optimal use of differences between internal representations of different stimuli.

It is well known that the processing of sound in the auditory system is accompanied by a loss of information,
because not every change in a stimulus can be detected. In intensity discrimination tasks using long-duration stimuli, the just detectable increment is approximately proportional to the pedestal level, in accordance with Weber's law. This implies that the auditory system is a nonlinear system, because otherwise the superposition principle holds and the level of a deterministic signal was of no relevance for the detection of a superimposed signal.

In fact, Weber's law has often been related to a logarithmic transformation of the linear pressure amplitude in the auditory system. Although this relation is not entirely consistent with some intensity discrimination data at high frequencies (Florentine et al., 1987; Long and Cullen, 1985; Carlyon and Moore, 1984 and 1986; Oxenham and Moore, 1995), it does provide a first-order approximation. But because the logarithm is strictly monotonic, this transformation alone is not equivalent to a loss of information. To cause a loss of information, a limitation of the resolution has to follow the nonlinearity. In the present model, such a limitation is achieved by adding an internal noise with a level-independent variance after the nonlinear transformation.

To simulate the ability of a human observer to discriminate between two auditory stimuli, an optimal detection process is attached to the model after the preprocessing stages. In this stage, differences between the internal representations of the two stimuli in question are analyzed. If the mean internal representations of the two stimuli differ by an amount that is large compared to the variance of the internal noise, the difference will be detectable with a probability close to unity.

In this way—if identical signals are presented by the same type of adaptive procedure as in corresponding measurements—the model could be considered as “imitating” a human observer. Hypotheses and models concerning “signal processing” in the auditory system can be tested by comparing simulated and measured data. The optimality of the detection process refers to the best possible theoretical performance in detecting signals under specific conditions. The mathematical framework for the optimal detection process was provided by the signal detection theory of Green and Swets (1966). In the following two sections, the successive stages of preprocessing of auditory stimuli and the general structure of the decision device are specified in detail.

B. Stages of peripheral processing

1. General structure

Because detection performance of the model depends on the preprocessing of the signals, it is necessary to describe the different processing stages in the auditory system and their realization in the model more precisely.

Sinusoids of different frequencies produce peak responses at different places along the basilar membrane (von Békésy, 1949). This frequency-place transformation implies that each place has a bandpass-filter characteristic, which was simulated by a linear basilar-membrane model (Schroeder, 1973; Strube, 1985). This model is realized as a wave-digital filter and computes the filtered signals in 120 output segments. Only the channel tuned to the signal frequency was further examined. As long as broadband noise maskers are used, the use of off-frequency information is not advantageous for the subjects.

The signal at the output of the specific basilar-membrane segment was half-wave rectified and low-pass filtered at 1 kHz. This stage roughly simulates the transformation of the mechanical oscillations of the basilar membrane into receptor potentials in the inner hair cells. The low-pass filtering essentially preserves the envelope of the signal for high carrier frequencies. Effects of adaptation were simulated by feedback loops (Püschel, 1988; Kohlrausch et al., 1992). This stage compresses stationary signals almost logarithmically whereas fast fluctuations of the input are transformed more linearly. These feedback loops were investigated to explain temporal masking effects and they will be described more precisely in the following subsection. In the stage following the feedback loops, the signal was low-pass filtered at...
8 Hz, corresponding to a time constant of 20 ms. This value was suggested by the outcome of simulations described in the accompanying paper (Dau et al., 1996).

2. Nonlinear adaptation

Since the nonlinear adaptation model of Püschel (1988) combined with the optimal detector process has an important influence on the results, this processing stage will be discussed in detail. The model tries to incorporate the adaptive properties of the auditory periphery (Kohrausch and Püschel, 1988; Kohrausch et al., 1992). The adaptation stage in the present model consists of a chain of five feedback loops in series with different time constants, as shown in Fig. 2. Within each single element, the low-pass filtered output is fed back to form the denominator of the dividing element. The divisor is the momentary charging state of the low-pass filter, determining the attenuation applied to the input. The time constants are ranged between 5 and 500 ms.1

This model has the following properties: After the onset of stationary signals the low-pass filters are charged according to their time constants. Because the charging state of the low-pass filter enters as divisor, a greater charging causes a greater attenuation of the input signal. For stationary signals an input value \( I \) produces a value of \( O = \sqrt[7]{I} \) at the output of the first feedback loop (derived from the stationary condition \( I/O = O \)). For a chain of \( n \) loops we obtain an output of \( O = \sqrt[2n]{I} \). For \( n=5 \) this approaches a logarithmic transform.

The range of derived output values for stationary inputs with levels between 0 and 100 dB was linearly mapped to the range of 0 to 100 model units (MU). For a perfect log transformation, a change of 1 MU would correspond to a 1-dB change at the input, irrespective of the input level. With five feedback loops we obtain an input–output characteristic, i.e., a transformation of the input signal level (in dB SPL) into model units (MU), as shown in Fig. 3. There is a small deviation from a straight line which would represent a logarithm transform on this scale. In the case of an input signal at a high level, since we assume a constant sensitivity in terms of output values (MU), this slight deviation from a logarithmic transform predicts a smaller just-noticeable change in level at high input levels compared with low input levels.

Input variations that are rapid compared with the time constants of the low-pass filters are transformed linearly. If these changes are slow enough to be followed by the charging state of the capacitor, the attenuation gain is also changed. Thus each element combines a static compressive nonlinearity with a higher sensitivity for fast temporal variations.

When a signal is turned off the charge of the capacitor, and therefore the attenuation applied to the input, decreases in each stage according to its time constant. Thus the time constants determine how fast the system returns to its state of rest. The charge of each low-pass element decays exponentially, which implies a linear decay on a dB scale. By combining several linear decays with different slopes, we get a piecewise approximation to the well known forward-masking curve (Hanna et al., 1982; Zwicker, 1984).

In this way a given signal can reduce the response to a signal presented subsequently. This makes it possible to simulate the nonlinear dependence of forward masked threshold functions on masker level and masker duration. With a short masker the stages with the long time constants cannot be fully charged and therefore only the decay due to the fast stages influences the forward-masked thresholds (cf. Kohrausch and Fassell, 1996).

To establish an absolute threshold for an auditory signal, the minimal value at the input of the adaptation stage is limited by a constant value. This can be interpreted as simulating “physiological noise” (e.g., Soderquist and Lindsey, 1972) that is added before the logarithmic transformation. If the current output of the basilar-membrane-filtered signal is smaller than the threshold it is replaced by the threshold value. In simultaneous masking situations this scaling is of little importance. When simulating mainly at absolute thresholds a stochastic noise instead of our constant threshold value should be used at this point in the model.

The calibration of the model is based on the 1-dB criterion in intensity discrimination tasks. In the first step of adjusting the model parameters, this value of a just-noticeable
change in level of 1 dB was used to determine the variance of the internal noise. A long-duration signal with a fixed frequency and a level of 60 dB SPL was presented as the signal. The variance of the internal noise was adjusted so that the adaptive procedure led to an increment threshold of approximately 1 dB. This variance of the internal noise was kept constant for all model simulations presented in the accompanying paper. Because of the almost logarithmic compression of signal amplitude in the model (cf. Fig. 3) the 1-dB criterion is also approximately satisfied over the whole input level range. The actual values for the just-noticeable change in level for input levels of 20, 40, 60, 80, and 100 dB are 1.44, 1.13, 0.95, 0.82, and 0.80 dB, respectively.

C. Structure of the decision device

In many psychoacoustical tests the observer’s task is to detect just-noticeable differences in a sound rather than the mere presence of a signal. At the beginning of a forced-choice masking experiment, for example, the signal is usually presented at a highly detectable level. This gives the subject a clear idea of “what to listen for.” We assume that the subject is able to get an “image” of the signal by comparing the suprathreshold signal in the masker with the masker alone. We assume, therefore, that the subject has access to the internal representations of the suprathreshold signal and the masker alone and that, in the run of an experiment, the subject uses these representations in deciding whether or not the signal was presented in a given interval.

Within the model, to get a representation of the signal, or of a change in a stimulus that is to be detected, the difference between the internal representation of the masker plus suprathreshold signal and the internal representation of the masker alone is considered. In Fig. 1 these representations are denoted by \( R(MT) \) and \( R(M) \), respectively. The difference is normalized to unity energy; this normalized signal is termed the template. It is important to note that the template is not derived directly from the internal representation of the signal alone, but from its representation against the masker background. This distinction is relevant for two reasons: First of all, the approach corresponds to the situation in an actual masking experiment, where the signal is typically not presented in isolation. Secondly, due to the nonlinear processing, the presence of the masker influences the internal representation of the signal in a nonlinear way.

A visual example of how the template is derived from internal representations of a suprathreshold signal and a masker alone is shown in Fig. 7, which will be described in detail in Sec. 1D. The top panel shows the temporal course of the internal representation of a masker that lasts from 100 to 300 ms. The middle panel shows the representation of the masker plus the supra-threshold signal, which is a short pulse occurring at 200 ms. The bottom panel shows the template, derived by subtracting the signals from the top and the middle panel and normalizing the result. The template was derived for the frequency channel that was tuned to the signal frequency. As long as broadband noise maskers are used, the use of off-frequency information is not advantageous in detecting the signal. The template can, however, without any modification be equally well derived from more than one peripheral channel. Such a multichannel template allows for off-frequency listening and spectral integration of signal energy across more than one critical band.

For the detection process, the model works on the difference between the internal representation of masker alone \( R(M) \) and of the current stimulus. If the test signal is within the current stimulus, given then by \( R(MTc) \), the difference signal \( [R(MTc) - R(M)] \) for deterministic stimuli contains only the change introduced by the signal, embedded in internal noise. If stochastic stimuli are applied, the difference signal is affected by the statistical properties of the stimuli in addition to the internal noise. As part of the detection process, the model computes the correlation between the template and the difference signal.

As it is shown in the Appendix, the correlation between the expected signal (template) and the signal actually received is a monotonic function of the likelihood ratio. It is known from signal detection theory (Green and Swets, 1966), that each decision rule that is based on the likelihood ratio, or another quantity which is monotonically related to the likelihood ratio, optimizes the decision maker’s goal (for example the number of correct decisions). In this sense the detection algorithm based on the correlation is a realization of an “optimal detector.”

In the simulations the stimuli were presented in the same way as in real psychoacoustical measurements (cf. Dau et al., 1996). Thresholds were determined with a 3IFC procedure with adaptive level tracking (two-down one-up), which tracks the point on the psychometric function corresponding to 70.7% correct (Levitt, 1971). In our simulations the stored reference, i.e., the internal representation of the masker alone, \( R(M) \), was subtracted in each interval so that, for deterministic signals, there were two intervals containing internal noise only and one interval containing internal noise and the signal. Then, in each interval, the correlation with the template as described above was computed. The simulated observer selected the largest value as indicating the signal interval. The decision was regarded as an “incorrect response” if that value stemmed from one of the masker alone intervals (see Appendix).

Instead of actually adding the Gaussian noise to the internal representations, the calculations can be performed in a less time consuming way. If the variance of the internal noise is adjusted in the way described in the previous section, we can from the internal representation for a certain signal value directly compute the probability of being correct with respect to the actual \( d' \) within each trial. In a 3IFC procedure, the signal thresholds of 70.7% correct corresponds to a \( d' \) value of about 1.26 (see the Appendix and also MacMillan and Creelman, 1991).

If only the mean values of the simulated thresholds are of interest, it is possible to simplify the procedure even further. In this case the model doesn’t behave truly randomly concerning the computed probability of correct decisions. Instead, the stimulus is assumed to be detected if the probability of being correct is larger than the expected value of the simulated measurement procedure (for example 70.7% for a two-down one-up algorithm). Otherwise, the stimulus is not detected. Therefore, we simply iterated the procedure to de-
termine the simulated thresholds. In such a simulation the psychometric function has a stepped shape. As long as deterministic signals are applied, the model provides deterministic mean thresholds. However, if signals with stochastic properties like random-noise maskers are used, the model also shows stochastic behaviour.

D. Example of auditory signal processing in the model

As an example of auditory preprocessing we will examine a section of a broadband noise masker with a flat power spectrum between 20 and 5000 Hz and a duration of 200 ms. Figures 4 and 5 show the time function and the power spectrum, respectively. The masker was filtered using the basilar-membrane model of Strube (1985). Only the filter output tuned to a signal frequency of 3 kHz was considered. The logarithmic power spectrum of this filter output is shown in Fig. 6. The top panel in Fig. 7 shows the internal representation which results after half-wave rectification, low-pass filtering at 1 kHz, adaptation by the feedback loops and the final low-pass filter (8 Hz). The ordinate is scaled in model units (MU). A pronounced overshoot occurs when the masker is turned on. The system adapts to its steady state before the masker is switched off. After the masker has been switched off, an undershoot is first observed, after which the system slowly returns to its state of rest. The middle panel in Fig. 7 shows the internal representation of the sum of the masker and a test signal well above threshold, \( R(MT) \). The test signal is a Hanning-windowed 10-ms 3-kHz sinusoid, which has been added to the masker of Fig. 4 with an onset delay of 100 ms. The onset of the test signal produces an overshoot of the internal "excitation" while the system is still in the process of adapting to the masker. Finally, the normalized difference between the two representations is shown in the bottom panel of Fig. 7. It is assumed that the

FIG. 4. Temporal waveform of a 200-ms noise masker.

FIG. 5. Logarithmic power spectrum of the masker of Fig. 4.

FIG. 6. Spectrum of the masker filtered by the basilar-membrane model of Strube (1985). Only the output of the filter tuned to a signal frequency of 3 kHz is plotted.

FIG. 7. Top panel: Internal representation of the masker (without internal noise). Middle panel: Internal representation of the masker plus test signal (without internal noise). The test signal was a Hanning windowed 10-ms 3-kHz sinusoid that was added to the temporal center of the masker waveform of Fig. 4. Bottom panel: Normalized difference between the internal representations of the masker plus test signal and the masker alone. This difference is called the template.
II. DISCUSSION

In the previous section, we have described a specific implementation of a model to predict masked thresholds. This model differs in certain respects from other approaches. The model proposed by Oxenham and Moore (1994) is designed to describe thresholds in forward, backward and combined forward- and backward-masking conditions. Following a concept proposed by Penner (1980), their data were modeled by subjecting the amplitude of the stimuli to a power-law nonlinearity followed by a sliding temporal integrator ("window"). They assumed that the threshold corresponds to a fixed signal-to-noise ratio at the output of the window. Their decision device is based on the largest difference in output between the interval containing the masker and signal, and the interval containing only the masker. Similar to the approach in the present study, such a decision mechanism requires that the subject retains a dynamic pattern of the internal representation of the masker alone. The model is able to account for the decay of nonsimultaneous masking. The temporal windows derived from the data are also able to predict thresholds in decrement and increment detection tasks, and to account for the longer-term effects of masker duration in forward masking. However, it is unclear, how well the model is able to predict thresholds in simultaneous masking and whether it can show differences for stochastic and deterministic maskers.

One difference between their model and the present one is the decision device. Oxenham and Moore used the output of the temporal window only at the point where the signal-to-noise ratio is greatest, so that there is no possibility of combining information over time. The authors noted, however, that a decision device based on information across the duration of the signal may have provided a better fit to some of their data. In the same vein, Viemeister and Wakefield (1991) have suggested that data from duration and intensity trade-offs may be accounted for by a "multiple-look" theory, which also involves combining information across time, such as is implemented in the present model.

Other temporal processing models have been proposed to account for various phenomena involving temporal resolution such as modulation detection (Viemeister, 1979), gap detection (Forrest and Green, 1987) and certain temporal aspects of nonsimultaneous masking (Moore et al., 1988). The models incorporate predetection bandpass filtering, followed by a nonlinearity, mostly half-wave rectification, followed by low-pass filtering. Forrest and Green proposed as a decision device the ratio of the largest to the smallest instantaneous output of the low-pass filter. The decision criterion used by Moore et al. is based on the largest instantaneous signal-to-noise ratio. Due to their short time constants and the fact the decision algorithms are based on a single point in time, neither model can describe temporal integration data.

Viemeister suggested as a decision statistic the coupled root-mean-squared (rms) value of the output of the low-pass filter calculated over the duration of the observation interval. The thresholds were defined as the modulation depth necessary to produce a certain average increment, relative to no modulation. Such a process also does not account for describing longer term integration data. Conversely, it is known that energy-based integration models, designed to account for longer term temporal integration (e.g., Green, 1960; Green and Swets, 1966), cannot account for temporal resolution data. Also, due to their long time constants such models are not able to predict differences in signal thresholds between running and frozen noise.

The model presented in this study has been initially designed to describe temporal aspects of masking. There is no principal restriction in its application with regard to duration, spectral composition and the statistical properties of the masker.

Across-channel processing, as is probably used in corepresentation masking release (CMR) and in binaural conditions, is not accounted for. This may be possible with an additional stage which calculates the correlation between the different channels.

As we use a linear basilar membrane filterbank the current model cannot predict nonlinear effects such as the level-dependent upward spread of masking, two-tone suppression, and combination tones. However, it can be expected that by replacing the initial filtering stage with a nonlinear, level-dependent basilar-membrane model these effects can be incorporated in a future version of this model.

III. SUMMARY

In this paper we have described a model that simulates the detection characteristics of the human auditory system. The model consists of several, partly nonlinear processing stages and an optimal detector as decision device. The idea of the optimal detector was taken from the signal detection theory of Green and Swets (1966). The statistical basis was provided by the assumption of an internal Gaussian noise following the nonlinear processing in the periphery of the auditory system. The nonlinear processing was developed to describe the main features of nonsimultaneous masking. The optimal detector makes it possible to check the influence of the preprocessing stages on calculated thresholds. The internal noise sets a limit to the detection of a change in acoustic stimuli and therefore determines the resolution of the auditory system. If identical signals and the same adaptive algorithm as in corresponding measurements are used, the computer may be regarded as an "artificial" observer. The standard deviation of the internal noise is determined in a basic simulation concerning the detection of a level increment. The five time constants of the adaptive nonlinear processing are determined in appropriate simulations concerning forward masking. These will be presented in the accompanying paper, in which some simulations and measurements of backward, simultaneous and forward masking will be discussed in detail.
ACKNOWLEDGMENTS

We would like to thank all our colleagues of the psychoacoustic research groups at the Universities of Göttingen and Oldenburg for fruitful discussions on the content of this paper. We also thank Dr. Brian Moore for his comments and suggestions concerning this study and Jesko Verhey, Andrew Oxenham, Martin Hansen, and Andres Sander for their critical reading of earlier versions of this paper. The two anonymous reviewers also provided very constructive criticism. This study was supported by a grant from the Deutsche Forschungsgemeinschaft (Ko 1033/3-1).

APPENDIX

In the statistical analysis it is assumed that, due to the internal noise, the representation $e(t)$ of a whole temporal (acoustical) stimulus at the time $t_n$ is given by the conditional density $P(e(t_n)|N) = P(e|N)$ if noise alone is present. It is characterized by $P(e(t_n)|S) = P(e|S)$ if signal plus noise are presented.

With the following simplifying assumptions:
1. $P(e|S)$ has mean $s_0$ and variance $\sigma^2$.
2. $P(e|N)$ has mean 0 and the same variance $\sigma^2$.

The probability of a noise-alone interval evoking a representation $e_{\nu}$ at time $t_n$ is given by

$$P(e|N) = \frac{1}{\sigma \sqrt{2 \pi}} e^{-\left(s_{\nu} - s_0\right)^2/2\sigma^2}. \quad (A1)$$

Given signal plus noise, the density function for the occurrence of the representation $e_{\nu}$ is denoted by

$$P(e|S) = \frac{1}{\sigma \sqrt{2 \pi}} e^{-\left(s_{\nu} - s_0\right)^2/2\sigma^2}. \quad (A2)$$

On the assumption that the different samples $e_{\nu}$ are uncorrelated with each other, the representation of the whole temporal stimulus is given by the product of the sample densities, for example,

$$P(e|S) = \Pi \, P(e|S). \quad (A3)$$

Now the likelihood ratio can be computed. It maps a whole temporal interval onto a positive real number:

$$l(e(t)) = \frac{P(e|S)}{P(e|N)} = \frac{\Pi_n \left(1/\sigma \sqrt{2 \pi}\right) e^{-\left(s_{\nu} - s_0\right)^2/2\sigma^2}} {\Pi_n \left(1/\sigma \sqrt{2 \pi}\right) e^{-\left(s_{\nu} - s_0\right)^2/2\sigma^2}}. \quad (A4)$$

If we take the natural logarithm and simplify this we obtain

$$\ln l(e) = \frac{1}{\sigma^2} \left( \sum \left( e_{\nu} s_{\nu} - \frac{1}{2} \sum \left( s_{\nu} \right)^2 \right) \right). \quad (A5)$$

According to Eq. (A5), the correlation between the expected signal $s(t)$ ("template") and the received $e(t)$ is a monotonic function of the likelihood ratio. According to the signal detection theory (Green and Swets, 1966) the correlation is therefore an appropriate quantity for expressing an "optimal" decision criterion.

To be able to compute the probability of correct responses we need to know the distributions of the likelihoods or correlations. If we regard the signals as vectors, the correlation is the scalar product of the Gaussian-distributed event $e(t)$ and the deterministic template $s(t)$, so the correlation must be Gaussian, too. If the template energy is normalized to 1, the distributions $P(l|S)$ and $P(l|N)$ have the same variance $\sigma^2$ as the distributions $P(e|S)$ and $P(e|N)$.

The mean of a noise interval is 0, the mean of the signal interval is denoted by $d$. The probability of being correct in an $m$IFC task is given by

$$P_m(\text{correct}) = \int_{-\infty}^{\infty} \phi \left( x - \frac{d}{\sigma} \right) \Phi(x)^{m-1} \, dx, \quad (A6)$$

with

$$\Phi(x) = \int_{-\infty}^{x} \phi(x') \, dx'. \quad (A7)$$

where $\phi(x)$ denotes the density of the standard normal distribution.

Therefore, according to Eq. (A6), the probability of being correct depends on $d$, $\sigma^2$ and the number $m$ of the observation intervals.

An alternative way of deriving $P_m(\text{correct})$ in Eq. (A6) is by computing the probability of the correlation value associated with the signal being greater than the maximum of all $m-1$ correlation values associated with noise alone. Cramer (1946) showed that the maximum of the noise samples has approximately a Gaussian distribution. Since the difference of two Gaussian "random variables" also has a Gaussian distribution, the difference between the sample associated with the signal and the greatest noise sample must also have a Gaussian distribution. Therefore Eq. (A6) can be approximated by a Gaussian integral with a mean $\mu^*(m)$ and a variance $\sigma^2(m)^2$ (tabulated in Green and Dai, 1991):

$$P_m(\text{correct}) \approx \int_{-\infty}^{d - \mu^*(m)/\sigma^*(m)} \phi(x) \, dx = \Phi \left( \frac{d - \mu^*(m)}{\sigma^*(m)} \right). \quad (A8)$$

In our simulations, this integral must be calculated whenever the value of the actual variable, which is in most cases the signal level, is changed. According to Press et al. (1989) a closely related mathematical integral, the so-called error function $\text{erf}(x)$, is easily computed

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_{0}^{x} e^{-t^2} \, dt. \quad (A9)$$

1In the original version (Püschel, 1988) the time constants were linearly distributed between 5 and 500 ms. In the current implementation, one of the larger time constants was replaced by a smaller one. This led to a better agreement between measured and simulated data in forward masking conditions. The time constants are 5, 50, 129, 253, 500 ms.

2In the model version described in this paper we have, for simplicity, assumed that the template is derived once at a clearly suprathreshold signal value. An alternative, and probably more realistic, implementation would allow a recomputation of the template with a decreasing signal level. Such
an implementation allows for changes in the detection cue used by a subject, which we observe in measurements with deterministic stimuli, when the signal approaches its threshold. The consequences of such a dynamic variation of the template are currently being investigated.

Alternatively, the correlation of the template with the individual test stimuli could be computed without subtracting the representation of the masker alone. If the difference between the correlation of the template with the signal interval and that with the masker-alone interval is considered, the results of both processes are mathematically equivalent. However, it is important for the detection process that in forming the template, the masker alone is subtracted from the original suprathreshold signal. For reasons of parsimony, we have chosen to describe the detection process within the actual experiment in the same way.

One might ask, what is gained by first deriving a template and comparing it with the various intervals instead of comparing the intervals directly with each other. For instance, as one reviewer suggested, why could one not directly compare the internal representations of the signals from the experimental intervals? The difference between such an approach and the template approach lies in the amount of external variability, which could be computed without subtracting the representation of the masker from the signal interval and that with the masker-alone interval is considered, the results of both processes are mathematically equivalent. However, it is important for the detection process that in forming the template, the masker alone is subtracted from the original suprathreshold signal. For reasons of parsimony, we have chosen to describe the detection process within the actual experiment in the same way.