A KALMAN FILTER MODEL FOR SIGNAL ESTIMATION IN THE AUDITORY SYSTEM

by

Martin M. Hauger

Submitted in partial fulfilment of the requirements for the degree

Master of Engineering (Electronic Engineering)

in the

Faculty of Engineering, the Built Environment and Information Technology

UNIVERSITY OF PRETORIA

February 2005

A Kalman filter model for signal estimation in the auditory system

by
Martin Manfred Hauger
Study leader: Prof. Johan Hanekom
Department: Electrical, Electronic and Computer Engineering
Degree: M.Eng. (electronic)

Summary

Using a Kalman filter that contains a forward-predictive model of a relevant system, to predict the states of that system by means of an analysis-by-synthesis implementation in order to evade significant time delays incurred by feedback mechanisms was previously applied to the coordinated movement of limbs by means of the cerebellum. In this dissertation, the same concept was applied to the auditory system in order to investigate if such a concept is a universal neurophysiological method for correctly estimating a state in a quick and reliable way. To test this assumption an auditory system model and Kalman estimator were designed, where the Kalman filter contained a stochastically equivalent forward-predictive model of the complete auditory system model. The Kalman filter was used to estimate the power found in a particular band of the frequency spectrum and its performance in the mean-squared error sense was compared to that of a simple postsynaptic current decoding filter under various types of neural channel noise.

It was shown that the Kalman filter, containing a biologically plausible internal model could estimate the power better than a postsynaptic current decoding filter, proposed in the literature. When the just-noticeable difference in intensity discrimination, as reported in the literature, was compared to model-predictions, it was shown that a smaller mean-squared error results in the case of the designed auditory system model and Kalman estimator.

This suggests that the application of the Kalman filter concept is important as it provides a bridge between measured data and the auditory system model. It was concluded that a Kalman filter model containing a biologically plausible internal model can explain some characteristics of the signal processing of the auditory system. The research suggests that the principle of an estimator that contains an internal model could be a universal neurophysiological method for the correct estimation of a desired state.

***Keywords*- Auditory system, estimation, Kalman filtering, modelling, analysis-by-synthesis, internal model, intensity discrimination, postsynaptic current, neural code, neural channel model.**

'n Kalmanfiltermodel vir seinestimasie in die gehoorstelsel
deur
Martin Manfred Hauger
Studie leier: Prof. Johan Hanekom
Departement: Elektriese, Elektroniese en Rekenaar-Ingenieurswese
Graad: M.Eng. (elektroniese)

Opsomming

'n Kalmanfilter met 'n vorentoe-voorspellermodel van 'n gegewe stelsel om die toestande van die stelsel te voorspel deur 'n analise-deur-sintese tegniek, is voorheen gebruik in die konteks van koördinering van beweging van ledemate deur die cerebellum. In hierdie verhandeling is dieselfde beginsels toegepas op die gehoorstelsel om te ondersoek of dit 'n universele neurofisiologiese metode is vir vinnige en betroubare skatting van toestande. Om dit te toets is 'n model van die ouditiewe stelsel en 'n Kalmanskatter ontwikkel, met 'n Kalmanfilter wat 'n stogastiese-ekwivalente vorentoe-voorspellermodel van die gehoorstelsel bevat. Die Kalmanfilter is gebruik om die drywing in 'n gegewe band van die frekwensiespektrum te skat, en die werkverrigting in 'n gemiddelde-kwadraat fout sin is vergelyk met dié van 'n eenvoudige postsinaptiese stroom dekoderingsfilter met verskillende tipes ruis in die neurale kanaal.

Dit is aangetoon dat die Kalmanfilter met 'n biologies-realistiese interne model beter doen as die postsinaptiese stroom dekoderingsfilter wat in die literatuur voorgestel word. Die netwaarneembare verskil in intensiteitsdiskriminasie soos beskryf in die literatuur word vergelyk met modelvoorspellings en daar word aangetoon dat die gemiddelde-kwadraat fout kleiner is vir die ontwerpte gehoorstelselmodel en Kalmanskatter.

Dit dui aan dat toepassing van die Kalmanfilterkonsep belangrik is, deurdat dit 'n brug bou tussen gemete data en 'n gehoorstelselmodel. Die gevolgtrekking is dat die Kalmanfilter met 'n biologiese-realistiese interne model sommige van die eienskappe van die prosessering van seine deur die gehoorstelsel kan verklaar. Die navorsing dui daarop dat die beginsel van 'n skatter wat 'n interne model van die eksterne werklikheid bevat, 'n universele neurofisiologiese metode mag wees vir die korrekte skatting van 'n verlangde toestand of sein deur die brein.

***Sleutelwoorde*- Gehoorstelsel, estimasie, Kalmanfilter, modellering, analise-deur-sintese, interne model, intensiteitdiskriminasie, postsinaptiese stroom, neurale kode, neurale kanaal model.**

Acknowledgments

To my dear parents, friends and colleagues without whom this would never have been possible.

LIST OF ABBREVIATIONS

AP	Action Potential
AR	Autoregressive
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
CANS	Central Auditory Nervous System
CDF	Cumulative Distribution Function
CR	Characteristic Frequency
CN	Cochlear Nucleus
CNS	Central Nervous System
CoZeCs	Converted Zero Crossings
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
EKF	Extended Kalman Filter
FHN	FitzHugh-Nagumo
IF	Integrate-and-Fire
IKF	Iterated Kalman Filter
ISI	Interspike Interval
JND	Just-Noticeable Difference
JNDL	Just-Noticeable Difference in Level
LIF	Leaky Integrate-and-Fire
LKF	Linearised Kalman Filter
MA	Moving Average
MLE	Maximum Likelihood Estimation
MMSE	Minimum Mean-Squared Error
MSE	Mean-Squared Error
PSC	Postsynaptic Current
PSD	Power Spectral Density
PST	Poststimulus Time
RV	Random Variable
RZ	Real-Zero
RZC	Real-Zero Crossing
SID	System Identification

SPL	Sound Pressure Level
WSS	Wide Sense Stationary

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Neurons communicate by producing sequences of fixed-size electrical impulses called action potentials (AP) or spikes. Perceptions, decisions and ideas can all be encoded into trains of action potentials, but the basis of this coding scheme is still not well understood. Deciphering this code is one of the primary goals in experimental neuroscience, which is especially interesting to engineers since parallels exist to well-established communications and signal processing theories.

Neurobiological coding schemes are present in the visual and auditory systems, amongst others. To facilitate augmentation of damaged or defective parts thereof, an in-depth understanding into the functioning of these parts is required. This has sparked an interest to model such systems in terms of their physiology and has found widespread applications in cochlear and visual implants.

1.1 PROBLEM STATEMENT

A thorough understanding of how the different parts of the central nervous system (CNS) exploit neural spike patterns reliably and swiftly given relatively unreliable components such as neurons, still poses an interesting problem. Particularly in the auditory system, APs are utilised to transmit information to the central auditory nervous system (CANS) and are used in some form or another to produce sound perception.

Previously a concept called the *analysis-by-synthesis method* was successfully applied to a motor control problem where the cerebellum coordinates the movement of limbs (Miall, Weir, Wolpert and Stein, 1993; Wolpert, Ghahramani and Jordan, 1995). It was argued that a *forward-predictive model* is present in the cerebellum that predicts the position of the arm given a particular motor command and a noisy measurement fed back by the visual system or proprioception.

The problem addressed was to investigate how neural coding is implemented in the central nervous system by taking an analysis-by-synthesis design approach rather than the traditional analysis approach. This was done by designing a coding/decoding system taking limitations of the auditory system into account. The biological implications of such a design were then investigated. The objective was to use this approach to improve understanding of neural coding.

1.2 NEUROBIOLOGY: AN ENGINEERING PERSPECTIVE

The human body or any biological organism for that matter contains a vast number of interacting systems making life as it is possible. It is not surprising that these intricate systems have caught the imagination of many scientists and engineers who found a need to explore them. An interest in the auditory system, being one of the most complex biological interfaces between a human being or animal and its surroundings, has been the catalyst in creating a community of people dedicated to gain an understanding into its highly complex nature. The insights gained have led to considerable advances made in the field of cochlear implants (Meyer-Bäse and Scheich, 1997; Loizou, 1999) and visual implants (Meyer, 2002; Humayun, Weiland, Fujii, Greenberg, Williamson, Little, Mech, Cimmarusti, Boemel, Dagnelie and de Juan Jr., 2003).

Any sensory system can be broken down into various stages of signal processing blocks, much like a communications system, the aim being to communicate a particular message from one point to another, reliably and effectively under the influences of the surrounding environment. An engineer can thus deduce that similar principles that apply to a communication system will also be applicable to such a biological system. One way of understanding the intricate details of the auditory system would be to analyse such a system with the help of properly designed and conducted experiments. The experimental results could then be employed to design a model of the processes, consisting of a series of black boxes each performing the function of the various steps found within the auditory system. This would result in a model producing outputs analogous to those of the auditory system.

To design such a model certain characteristics of the auditory system can guide one, something that is often neglected. Specifically, the auditory system is a dynamic system as the behaviour of the outer haircells suggest (Yost, 2000). Depending on the type of stimuli received by the stereocilia of the outer haircells through the cochlea, the stereocilia can stiffen or relax resulting in variable frequency selectivity, due to mechanical coupling between the basilar and tectorial membrane (see *Chapter 2*). There however seems to be some form of feedback mechanism both locally and via the CANS (Warren, 1999) that connects the inner haircells and the outer haircells. Cochlear emissions (ringing) (Yost, 2000) observed 5-10 ms after delivering a click to the cochlea, where ringing is induced by outer haircell motility, is a good example of a local form of feedback. Such short reaction times however do not seem biologically plausible when transmission delays between the inner haircells, the central audi-

tory nervous system (CANS) and the outer haircells are taken into account. It thus seems plausible that some form of system or mechanism exists within the auditory system that can circumvent the delays incurred by feedback loops by predicting the measured information fed back.

A similar argument was brought forward by Liberman and Mattingly (1985) and Miall et al. (1993). In the coordinated movement of limbs by means of the cerebellum, it was argued that any feedback system involving the eye as a sensor would result in significant delays. In fact just the feedback loop from retina to motor cortex would introduce a delay of 200-250 ms. Feedback loops however cannot explain the fast reaction times found in human movement controlled via the cerebellum, where an entire movement takes place in 200-300 ms. This suggests that some form of an internal model that can predict the outcome given a particular neuronal command, could be present.

By using a forward-predictive model as shown in *Figure 1.1* (Miall et al., 1993) an appropriate action can be estimated and used before sensory feedback is available.

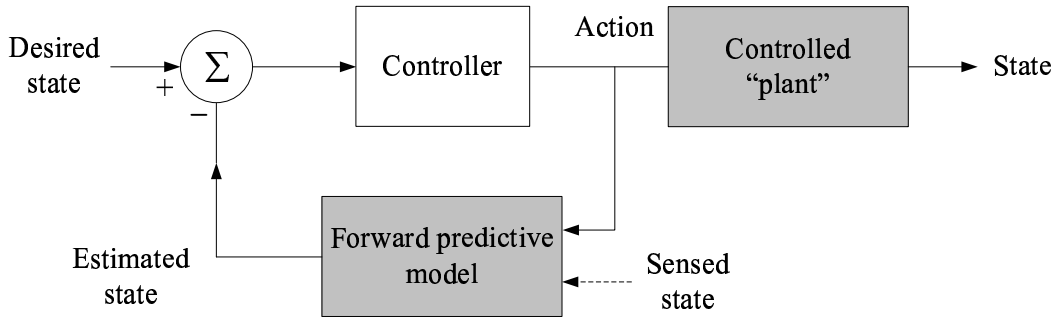

Figure 1.1: Control strategy using a forward-predictive model in negative feedback loop (Adapted from Miall et al. (1993)).

A forward-predictive model, predicting the outcome to a particular control stimulus could provide an estimate to what the outcome would be before the feedback network of sensors would report the result. This would result in the initiation of corrective measures in a very short time. The CNS however cannot know the real outcome of the action until actual feedback is received and ignoring it altogether would result in instability and then total collapse of the control system if only a slight error in the feedback predictive model were present. These errors thus need to be incorporated in some way into the model.

Wolpert et al. (1995) did just that, by applying a Kalman filter to the problem. By taking feedback measurements into account, the estimator corrects any errors incurred by the forward-predictive model.

In *Figure 1.2* the Kalman filter model for the motor system process is given.

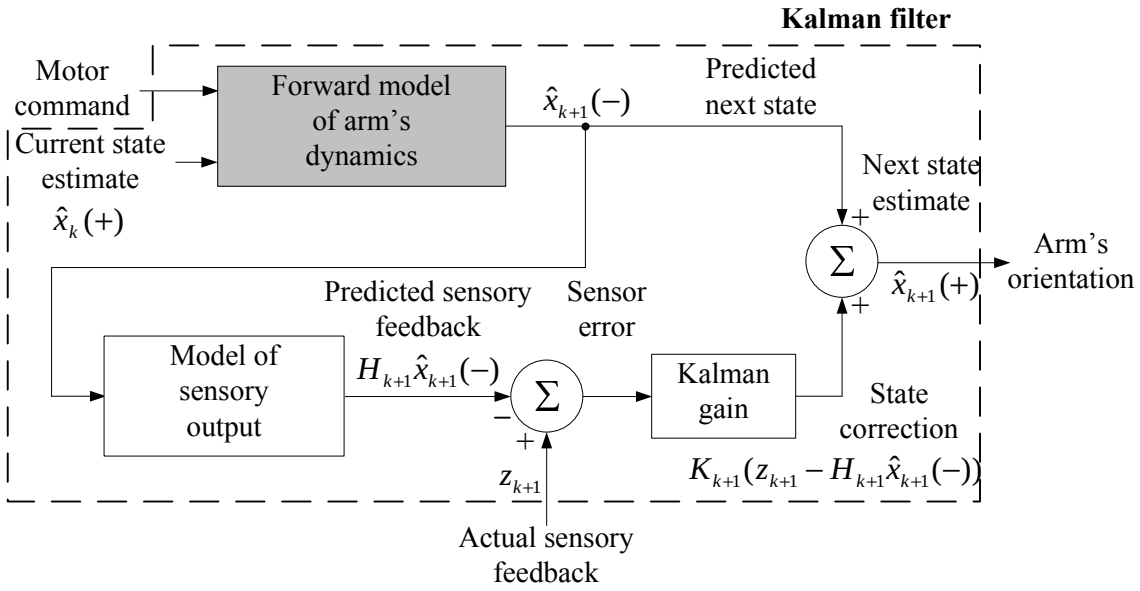

Figure 1.2: Kalman filter model for motor system process. The internal model is comprised of the forward model of the arm's dynamics and the model of the sensory output (Adapted from Wolpert et al. (1995)).

Motor commands given by the cerebellum are, together with the current state of the arm, processed by an internal model of the arm's dynamics. The predicted next state is also used to drive a sensory model of which the predicted sensory feedback is then compared to the actual sensory feedback. In this particular case, a Kalman filter based on the observer framework is used to adjust the predicted next state in order to find the next state estimate. The Kalman filter has access to *both* the motor command generated by the cerebellum (controller action in *Figure 1.1*) and sensed state of the arm via the visual feedback system. The bias and variance of experiments conducted and simulations performed by Wolpert et al. (1995) were compared and shown to be closely related, suggesting that an analysis-by-synthesis method using a Kalman filter can be used to describe the motor control process. The only problem that exists with using the Kalman filter configuration depicted in *Figure 1.2* is the fact that the system is now subjected to delays incurred by the sensory feedback loop via the visual system as was the case in *Figure 1.1*.

Nevertheless, similarly to the proposition made by Wolpert et al. (1995) it could be suggested that the auditory system contains an internal model of the auditory system in the form of a forward-predictive model and an estimator. It was shown (Tobey, 1993) that hearing is fundamental to a persons ability to speak. For children that learn to speak, hearing provides a feedback that is used to improve their speech-production skills. After one has learned how to speak, the internal model as such has been "calibrated" for a particular language. In fact by slightly adapting *Figure 1.2* a similar estimator model can be created for controlling speech generation by means of feedback via the auditory system.

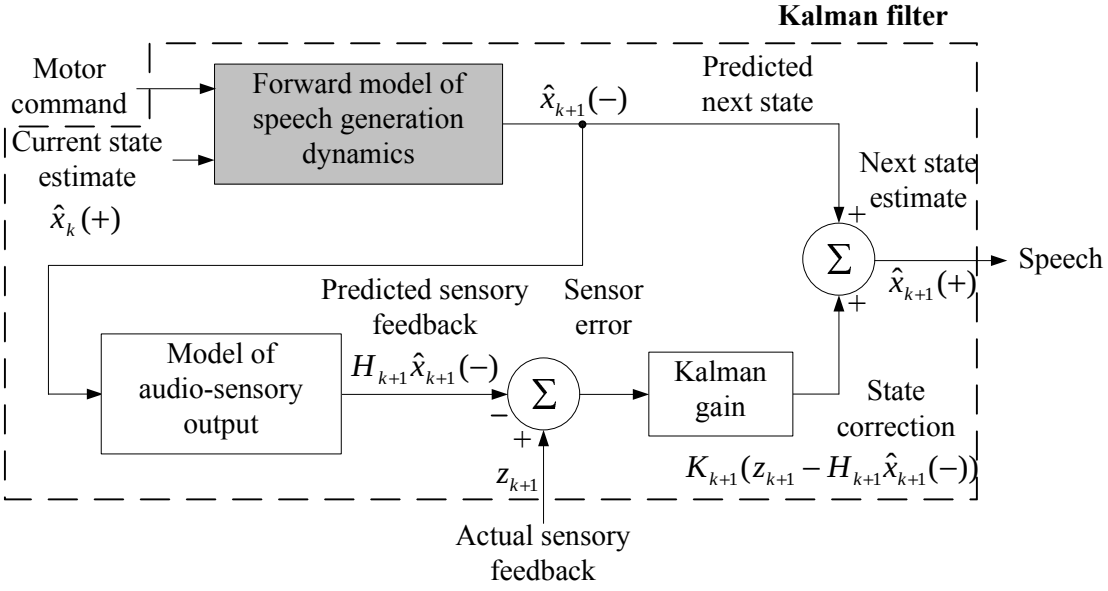

Figure 1.3: Kalman filter model for speech process.

In *Figure 1.3* the forward-predictive model for the arm is replaced by a model that converts a motor command created by the CNS to voiced and unvoiced speech sounds (see *Chapter 4*). The sensory mechanism in this case is not the visual system but the auditory system.

Supporting the proposition is the following observation made by Blamey, Arndt, Bergeron, Bredberg, Brimacombe and et al. (1996) and documented by Loizou (1999). The factors affecting the performance of cochlear implant patients are amongst others:

1. The duration of deafness,
2. age at onset of deafness and
3. duration of cochlear implant use.

Taking into account these factors, Blamey et al. (1996) developed a three-stage model of auditory performance for postlingually deafened adults as shown in *Figure 1.4*.

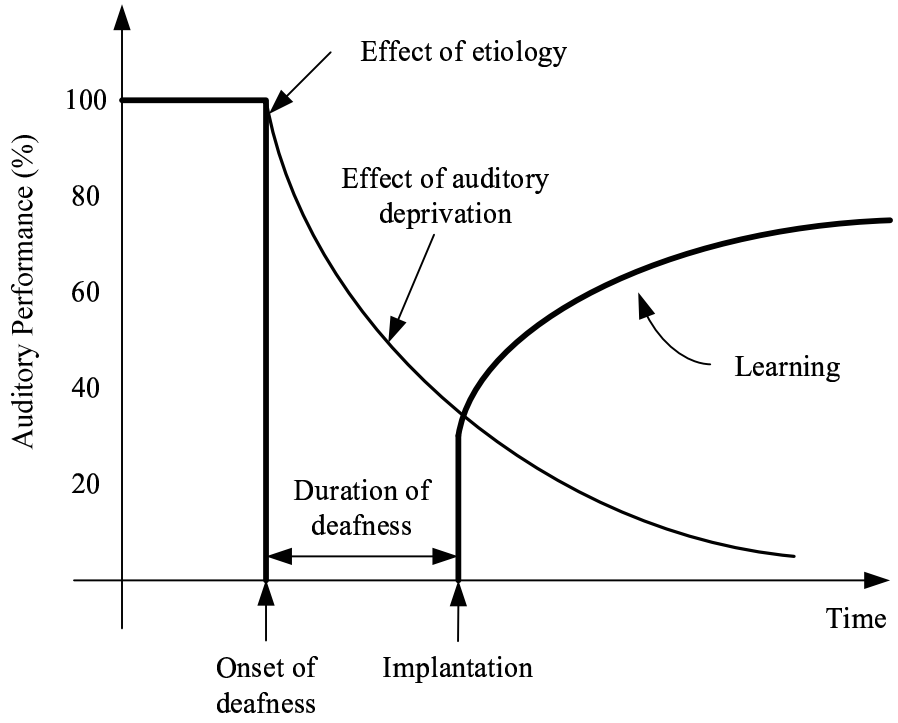

Figure 1.4: A three-stage model of auditory performance for postlingually deafened adults (Louizou 1999). The thick lines show measurable auditory performance, and the thin line shows potential auditory performance.

Stage 1 begins after normal language development where the auditory performance is close to 100 %. The auditory performance is defined here as the ability to discriminate, detect, identify or recognise speech. Stage 2 begins at the onset of deafness and decreases are dependent on the etiology of the hearing loss as well as continued deprivation of sound. Stage 3 begins with implantation and the patient immediately attains improvement in auditory performance, depending on the duration of deafness. As the patient's experience with the implant increases, the level of auditory performance rises. Seeing that haircells that are severely damaged in mammals cannot recover or be replaced (Yost, 2000) the only true gains achieved in terms of auditory performance must be attributed to some form of learning. Learning thus would justify the use of some form of an internal model that learns to mimic the behaviour of the auditory system and by doing so, takes on an assisting role.

It has been shown as early as 1980 by Subtelny (1980) that it was possible to teach deaf subjects how to speak in a intelligible way but that deviations in the speech production was

present, most notably in the speech rhythms. Cooper (1991) also documented how a combination of speech therapy combined with the implantation of a cochlear implant increased the overall speech production performance as far as melody, rhythm and intonation was concerned. Seifert, Oswald, Burns, Vischer, Kompis and Haeusler (2002) did a study with cochlear implant subjects and compared the formant frequency positions between prelingually deafened children and children with normal speech development. No significant difference in the fundamental frequency was observed but broader scattering of the higher formant frequencies f_1 to f_3 was observed in subjects with implantations done after the age of four years. This would indicate that learning correct speech production as far as the formant frequencies are concerned becomes increasingly difficult after the age of four. This is a good indication that some form of *trainable* internal model for speech production is present that can be taught by means of a feedback mechanism via the auditory system and good speech therapy.

Most interestingly however was a rare study done by Ito, Suzuki, Toma, Shiroma and Kaga (2002) in which a 6-year-old boy lost his normally-developed language ability within 2 months after bilateral sudden peripheral deafness. The patients *expressive language age*, a measure of normal development of language abilities in regards to age, regressed to a 9-11 months expressive language age. Within 10 months after the introduction of a cochlear implant the patient had reacquired most of his language abilities and at the age of 6 years and 9 months had a expressive language age of 5 years. It was estimated that the patient would within two years catch up with his peers as far as his language abilities were concerned.

It is evident from this study that an "internal model" as such has both the ability to deteriorate in performance once the feedback mechanism is interrupted as was evident in the 6-year-old boys regression period. It is also evident from the study that the model can be "recalibrated" once the feedback mechanism is restored. It is thus highly plausible that some form of internal model is present in the CANS, CNS or both which relies on a feedback mechanism.

It is thus interestingly enough to note that a Kalman filter is nothing more than a system that estimates states by using an internal model and measurements from the noisy process. When no measurements are fed back to it, the Kalman filter is forced to rely solely on its internal model. If noise is present, the state estimates start drifting from the true states resulting eventually in the divergence of the filter. After new measurements are introduced, the Kalman filter can in some cases eventually start tracking the true states again. It can thus be argued that a Kalman filter could assist the auditory system in signal estimation required

for reliable sound perception.

1.3 PRIMARY RESEARCH QUESTION AND HYPOTHESES

As shown, analysis-by-synthesis may provide useful insight in how the CNS successfully copes with possible errors induced and the delays incurred. In addition it provides a mean to design and test models. Having followed the logical steps as was done up to now the following question could be asked: *Does the presence of internal models as was discussed suggest that it is a universal neurophysiologic method for correctly estimating a state in a quick and reliable way? If so what are the biological implications thereof?* This is the primary research question.

To address this question an analysis-by-synthesis design approach, not to be confused with the analysis-by-synthesis implementation used by the Kalman filter, was taken to design a coding/decoding system that takes certain biological constraints into consideration. By implication the hypotheses thus are as follows:

1. There exists an internal model in the auditory system, that by means of an analysis-by-synthesis implementation, can explain the functioning of the system and
2. it is possible to take an analysis-by-synthesis design approach as compared to the more traditional analysis approach and by doing so to gain a better understanding of the auditory system and neural coding.

1.4 APPROACH

As was mentioned, the approach was to find an internal model by means of a design methodology that takes certain biological constraints into consideration. The primary objective was to investigate the biological implications that such an approach has, in particularly what signal processing steps are required and if the biology can support the implementation thereof. While the former objective was investigated extensively, the latter was not directly addressed in this dissertation. More detail on this approach is to follow.

The auditory system when receiving an acoustic signal in some form or another, which may be temporally or spectrally represented, needs to process the information in such a way that

the CANS can interpret this information correctly. The information that reaches the CANS is however not always reliable and according to the hypotheses depends on an internal model and an estimator to make the best estimate of what information was transmitted. The system that assists the auditory system in processing the information in the CANS could be similar to *Figure 1.2* and *Figure 1.3* and is depicted in *Figure 1.5*.

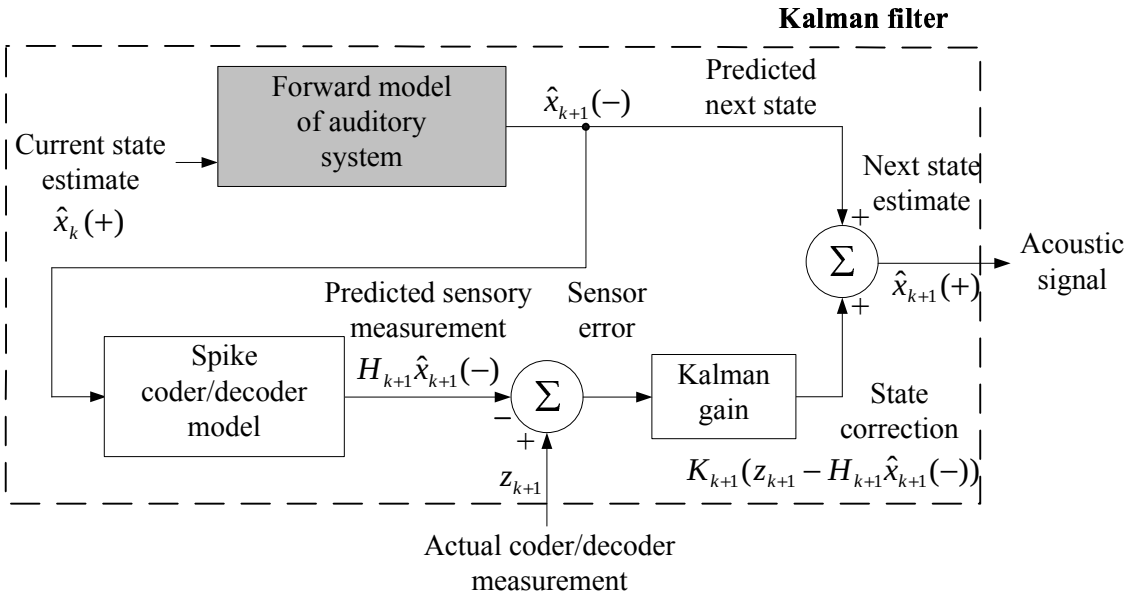

Figure 1.5: Kalman filter model for auditory system.

However, unlike the Kalman filter model used by Wolpert et al. (1995) and depicted in *Figure 1.2* the Kalman filter has only access to a stochastic model of the acoustic signal, found in the *forward-predictive model of auditory system* of *Figure 1.5*, rather than the true acoustic signal. The motivation for this is discussed in *Chapter 4*. Thus unlike the observer configuration used by Wolpert et al. (1995), the model in *Figure 1.5* only has access to the measurements taken after the actual spike coding/decoding takes place. These measurements are fed back to the Kalman filter as is indicated by the *actual coder/decoder measurement* label shown in *Figure 1.5*.

The signal from the acoustic source is processed by a forward-predictive model mimicking the behaviour of the auditory system. At its output the forward-predictive model contains information found in the mechanical vibrations of the basilar membrane just before being converted to APs by the inner haircells. At any given moment in time, the information is represented by states, which the Kalman filter needs to predict by observing the information through noisy measurements. The measurements of the information are however only made

after it has been coded to APs. It was assumed that the CANS has to decode the APs in order to interpret the information contained within the spike trains. The sensory model of *Figure 1.2* and *Figure 1.3* can be replaced by a spike coder/decoder model.

Before the Kalman filter can be designed and the Kalman gain calculated, a good model of the cochlea has to be found. This model serves as a reference point to the physiology of the auditory system.

1.4.1 Auditory System Modelling

The auditory system extends from the pinnae (see *Chapter 2*) up to the CANS. However, little is known about what happens after the inner haircells and cochlear nucleus (CN) (Eggermont, 2001), as far as interpretation of APs by the CANS is concerned. To avoid speculating about neural interpretation higher up in the brainstem this model only included process's up to and including the induction of spike action potentials.

In *Figure 1.6* a functional diagram depicts the auditory system.

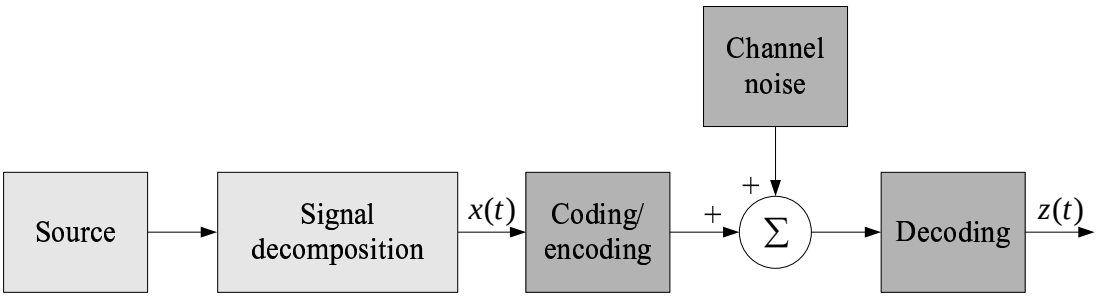

Figure 1.6: Model of auditory system.

A *source* signal, entirely defined by amplitude, frequency and phase at any given point in time, is broken down into smaller manageable blocks of data by the *signal decomposition* functional unit before the information contained within it is coded. The information is then transmitted by a neural channel, which is inherently noisy. The *decoding* functional unit is not essential in the auditory system as such seeing that little clarity exists on how exactly the CANS utilises the information coded in the APs. It was however included here, seeing that it was used as part of the testing mechanism to compare data generated by the auditory system model to that obtained by the estimator.

1.4.1.1 Signal decomposition and channel model

The travelling wave of the basilar membrane vibrates with a maximum amplitude at a position along the cochlea that is dependent on the frequency of stimulation (Yost, 2000). Each nerve fiber is most sensitive to a particular frequency and the individual fibers within the auditory nerve are also organised according to the particular frequency at which they are most sensitive and the location along the cochlea at which they innervate haircells. The basilar membrane thus acts like a bank of bandpass filters, which overlap to some extent in the frequency domain. Combined with the *tonotopic* nature of the cochlear nerve bundle, the mechanical to neural transduction can be thought of as a spectrum analyser spreading the information of the spectrum of the signal across a large number of channels, which are in this case a string of nerve fibers. But instead of coding the filtered signal in terms of amplitude, frequency and phase, the signal is demodulated with itself and only the power contained within a particular band is transmitted (*Chapter 4*). The power of the entire audible frequency spectrum is then carried among the unreliable auditory nerve fibers as will be discussed in *Chapter 5*.

1.4.2 Signal Recovery by Means of Estimation

Due to an unreliable transmission medium an estimator is applied to the auditory system in order to estimate the signal that was present just before it was coded by the inner haircells. The aim is to organise the signal power distribution among the auditory nerves in a map and to do so reliably, given a noisy channel (*Chapter 5*). No higher-level processing was applied to this power spectral map in order to avoid speculations on how the CANS interprets the information.

In order to apply an estimator such as the Kalman filter to the auditory system model (see *Figure 1.5*) a stochastically equivalent internal model subjected to various Kalman filtering restrains and conditions needs to be found. A short introduction to Kalman filters is given in the next section.

1.4.3 The Kalman Filter

A completely deterministic process with a plant and measurement process model can be evaluated uniquely and infinitely many times given a set of initial conditions seeing that no

degree of stochasticity exists. However, as soon as some randomness is introduced by, for example, noisy measurements, only an *estimate* can be found of a particular state of the stochastic system. The need thus arises to, in some optimal way, estimate state variables.

The Kalman filter being an optimal estimator in the minimum mean-squared error (MMSE) sense (see *Section 6.6*) can do just that. It estimates the conditional mean and covariance of the probability distribution of the state of a linear stochastic system with *uncorrelated Gaussian* process and measurement noise. The Kalman filter thus estimates states based only on the first two moments of the stochasticity seeing that uncorrelated Gaussian process and measurement noise can be uniquely defined by its mean and variance.

To apply this knowledge to the auditory system model given in *Figure 1.6* a stochastic equivalent model in the form of a recursive integrator model needs to be found, which is more commonly named a *shaping filter* (see *Subsection 6.7.1.1*). In *Figure 1.7* the *special* form of a shaping filter is depicted (see *Chapter 6* for general form).

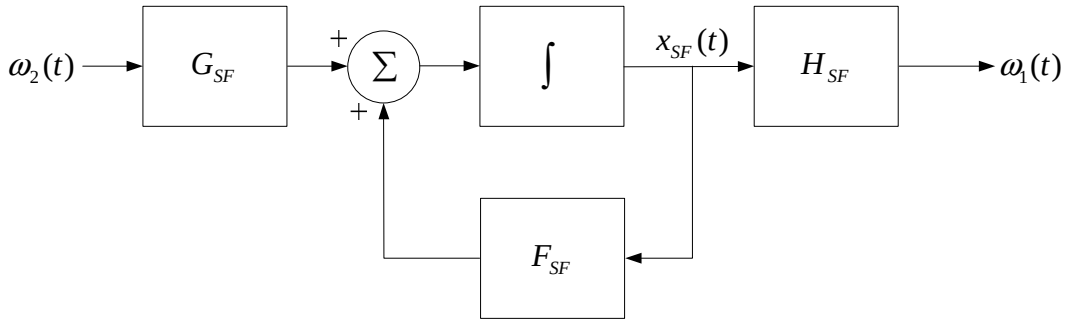

Figure 1.7: Special form of continuous-time shaping filter.

The uncorrelated Gaussian noise $\omega_2(t)$ multiplied by process noise coupling matrix G_{SF} is integrated to give the state $x_{SF}(t)$. The results is recursively fed back via the dynamic coefficient matrix F_{SF}. The state $x_{SF}(t)$ is also observed via the measurement matrix H_{SF}. Written as a set of equations the process noise can be written as

$$\dot{x}_{SF}(t) = F_{SF}(t)x_{SF}(t) + G_{SF}(t)\omega_2(t), \tag{1.1}$$

and the measurement equation as

$$\omega_1(t) = H_{SF}x_{SF}(t). \tag{1.2}$$

The auditory system in *Figure 1.6* is however not in the state-space form of *Equation (1.1)* and *Equation (1.2)*.

The motive behind finding a state-space equivalent in the first place if none is present is made clear by the continuous system model, measurement model and Kalman filter system diagram as depicted in *Figure 1.8*.

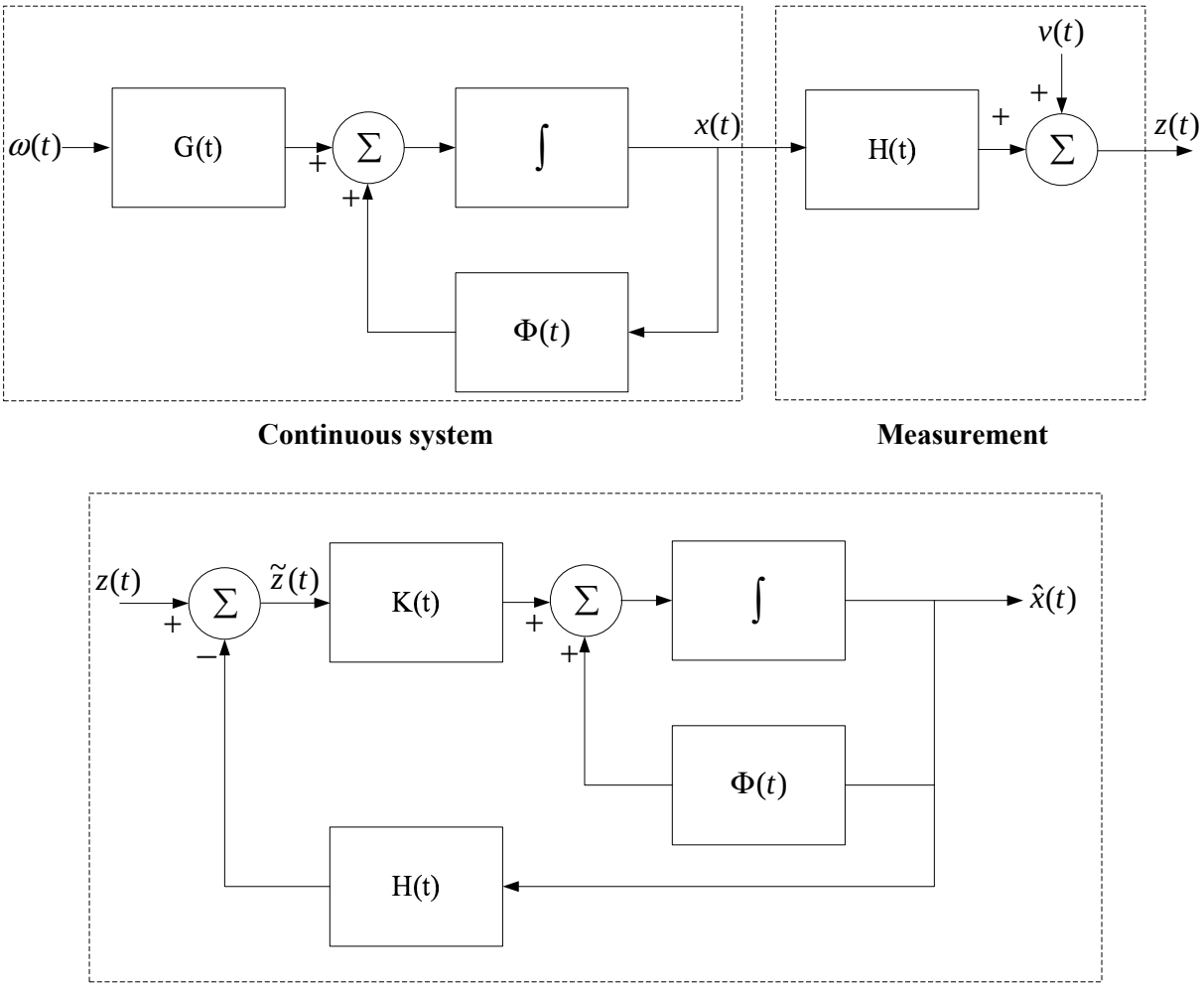

Figure 1.8: Continuous Kalman filter (bottom) with continuous system model and measurement model (top). The continuous system and measurement model represent the true process, while the continuous Kalman filter estimates the state variable based on an internal representation of the continuous system and measurement model. $\Phi(t)$ represents the forward-predictive model, H(t) the measurement or sensory model and K(t) the Kalman gain. $\Phi(t)$, H(t), K(t) and G(t) are all matrices.

Given a system, in this case continuous in time, and a measurement model, the continuous Kalman filter can recursively and *optimally* estimate the state variable $x(t)$ (*Chapter 6*). The Kalman filter assumes that the continuous system and measurement model can be modelled

as shown in *Figure 1.8*, where a white-noise source $\omega(t)$ drives a recursive linear process to generate a state variable $x(t)$ and a white-noise source $v(t)$ corrupts the measurements taken via the matrix H(t). It is evident from the diagram at the bottom of *Figure 1.8* that the "internal model" of the continuous Kalman filter contains information about the system and measurement models in the form of the matrices $\Phi(t)$ and H(t). By means of such an internal model or *whitening filter* the Kalman filter can optimally estimate the state variable $x(t)$. Thus if a process such as the auditory system can be mathematically described in terms of state-space equations and is included in the internal model of the Kalman filter, optimal estimation of the desired state variables is possible.

This is one of the main requirements of a Kalman filter and thus by implication the next task is to find state-space equations for the process given in *Figure 1.6* resulting in *Figure 1.9*b.

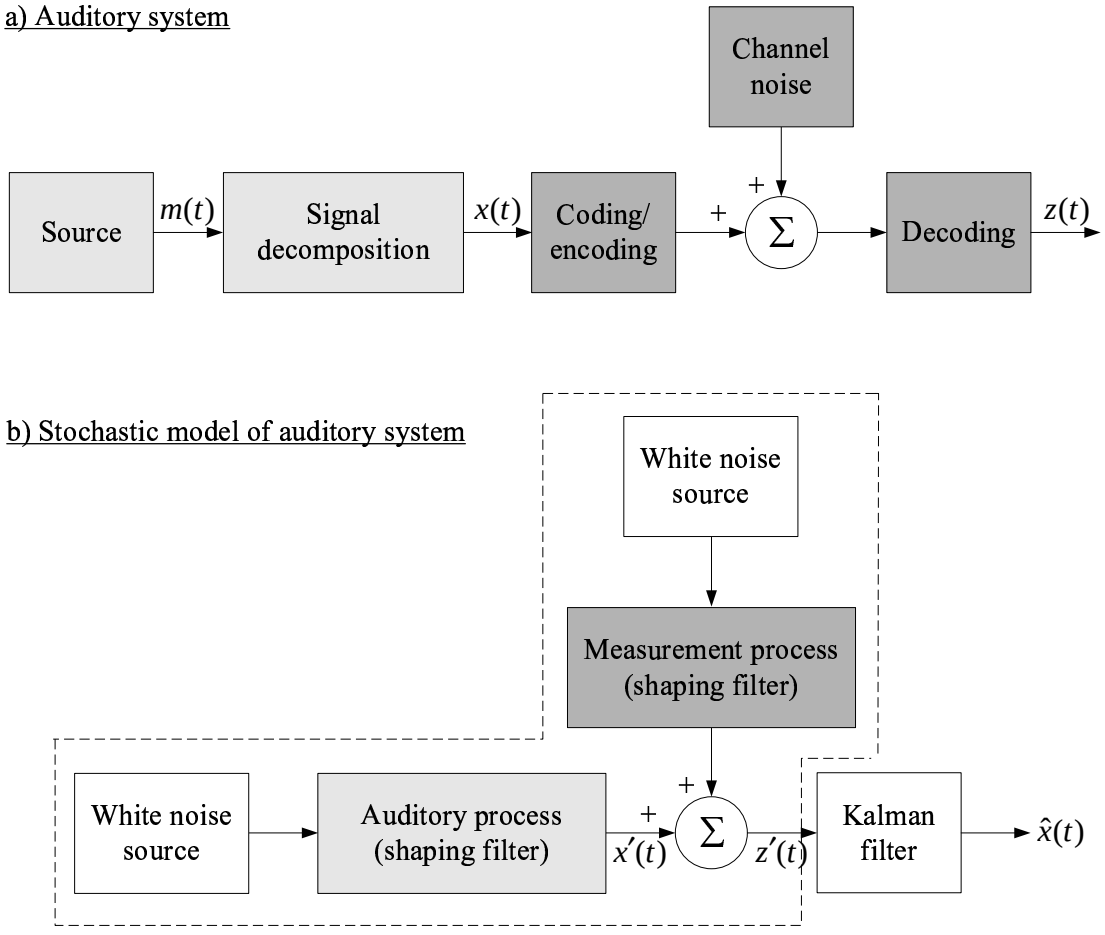

Figure 1.9: Auditory system model a) and stochastic model for a) shown in b). The Kalman filter contains an internal model of the stochastic model. $x'(t)$ and $z'(t)$ are the stochastic equivalents of $x(t)$ and $z(t)$.

The logical functional units in a) need to be replaced by equivalent state-space equations to form the estimator equivalent process in b). The *source* and *signal decomposition* blocks in a) can however be logically grouped in one large state-space equation to form the *auditory shaping* filter in b). Similarly the *encoding/coding*, *channel noise* and *decoding* functional unit can be modelled by the *measurement* shaping filter. The reason why the encoding/coding functional unit was not included in the auditory shaping filter will become apparent in *Chapter 6*.

Thus by rewriting the auditory system in terms of state-space equations a very similar estimator equivalent process can be found allowing for optimal estimation of the state variables as was already mentioned. To a reader familiar with Kalman filters it might already be evident that the process depicted in *Figure 1.5* is the Kalman filter found in *Figure 1.9*B and the *actual measurement feedback* in *Figure 1.5* is obtained from the auditory system model. Once the relevant Kalman filter has been found for the stochastic model of the auditory system, the Kalman filter can be applied to the auditory process in *Figure 1.9* to estimate the states $x(t)$ by observing the measurements $z(t)$.

One of the primary reasons why a Kalman filter is more relevant for this particular modelling problem shown in *Figure 1.9* is because a degree of stochasticity exists within the model itself. Even if a perfect model of the cochlea were to be found, additional noise sources would most certainly be present. Neurons for example (see *Chapter 5*) are inherently noisy. Thus to model the sensory functional unit in *Figure 1.5*, some form of stochasticity has to be introduced to the system. The Kalman filter is a suitable choice seeing that a combination of deterministic and stochastic input and measurement signals can be modelled.

Another advantages of the Kalman filter is its recursive implementation, which requires only the knowledge of the previous states to make an estimate thus greatly reducing memory requirements and increasing the computational efficiency.

1.4.4 Testing the Hypotheses

Once the Kalman filter has been found and applied to the auditory system the validity of the hypotheses can be verified. By evaluating the performance of such an estimator given an internal model linked closely to the physiology of the auditory system, it can be deduced if

the assumption that an internal model exists in the CANS is a plausible one. Furthermore the implications that such an internal model has on the biology can also be found.

1.5 OBJECTIVES OF DISSERTATION

The objectives of this dissertation were

1. to design and build a complete model of the auditory system up to and including the inner haircells,
2. to design and test an estimator containing a biologically plausible internal model in the form of a forward-predictive model and
3. to gain by means of the above mentioned *analysis-by-synthesis* method an understanding into the challenges the biology might face when trying to convey information using some form of a neural code.

The tools that were used to reach the above mentioned objectives were *modelling* and *simulation*, which are discussed in more detail below.

1.5.1 Modelling

A key part in understanding a particular process is to model it and by means of simulations, experiments and measurements validate the assumptions made. Although this approach provides models that mimic the *functionality* of a given process, they shed only little light on the *original concept* behind the process. The analysis-by-synthesis design approach was thus used to model a concept, given certain physiological limitations. In particular the auditory system was studied. The aim was to find a model that could extract information from any real signal and represent it in such a way that neural coding was straightforward.

The model, simulated in MATLAB provides an insight into the concept behind auditory signal processing and a mean to compare performances and the implications thereof. By doing so an in depth understanding was gained into the signal processing demands required to convey information in a meaningful yet very reliable way and how the limitations affect the process as a whole.

1.5.2 Estimation

Estimation methods provide a tool to extract information from a noisy environment, given the approximate knowledge of the underlying process. In this case, the estimator was required to make an estimate of the decoded signal at the output of the auditory system model (see *Figure 1.6*) given a noisy measurements. The argument was that the brain would have to be able to reliably estimate any form of information contained within the original signal. It was thus of paramount importance that the estimation process was as close to optimal as possible.

1.5.3 Note on Units Used

The units that are used in this work are picked relative to the leaky integrate-and-fire neuron in *Chapter 5* since its the only functional unit that is based on an electrical model with very specific units. Thus, units of volts are used throughout. These voltages are not actual internal voltages that can be measured. Rather, they relate to external sound pressure level (SPL) as explained in *Section 7.7*. The voltage units used also translate to a specific spike rate (which is measurable) as explained in *Subsection 5.5.2*.

1.6 REVIEW OF LITERATURE VS. CONTRIBUTION

Interest in the field of auditory perception has drawn the attention of the speech and hearing sciences, neurobiologists and engineers among others. Each has contributed a great deal towards the pool of knowledge that has been gathered over the last couple of decades. The focus in the different disciplines has however always been on only part of the auditory system.

The auditory system as was discussed earlier can be divided into logical functional units as is illustrated in *Figure 1.10*.

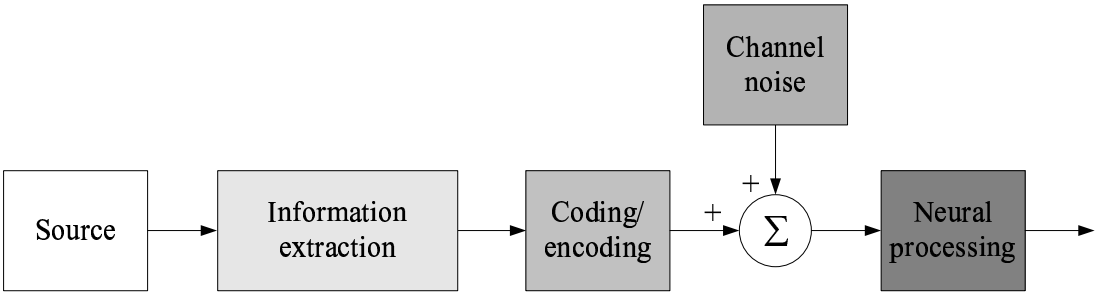

Figure 1.10: Auditory system given as a sequence of signal processing steps.

Some have combined information about most parts of the auditory system (Flanagan, 1972; Yost, 2000; Warren, 1999; Deutsch and Deutsch, 1993; Allen, 1985). However, due to the immense complexity of the auditory system as a whole, most application based research tends to focus on only parts of the auditory system. Models for the auditory system for example have been designed and discussed by Flanagan (1972), Deutsch and Deutsch (1993), Meyer-Bäse and Scheich (1997) and Doh-Suk, Lee and Rhee (1999). The focus of neurobiological research on the other hand lies more on the mechanical to neural transduction side, specifically the interpretation at neuronal level (Rieke, Warland, de Ruyter van Steveninck and Bialek, 1997; Eliasmith and Anderson, 2003; Deutsch and Deutsch, 1993; Maass and Bishop, 1999; Bialek, Rieke, de Ruyter van Steveninck and Warland, 1991; Abbott, 1994).

Others have applied Kalman filtering to auditory system modelling, to track either lower or higher hierarchial level information such as interspike intervals (ISI) (Hammarberg, 2002; Hanekom and Krüger, 2000; Gray, Slocumb and Elton, 1994) or formant positions and intensities (Niranjan and Cox, 1994; Lu and Doerschuk, 1996). The difficulties with tracking spike trains by either employing ISI tracking or doubly stochastic point processes (Snyder, 1975) is finding a suitable process and measurement noise model.

The first step to designing a Kalman filter for either lower or higher hierarchial level information tracking, is to design its process and measurement noise models as depicted in *Figure 1.11*.

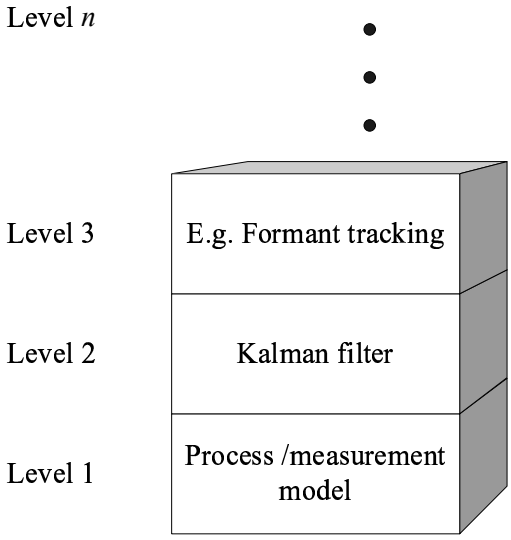

Figure 1.11: Kalman filtering applications for different hierarchial level problems.

The process and measurement models form the foundation with which the Kalman filter estimates state variables and solves higher level problems. All higher hierarchial level operations are solely dependent on the validity and correctness of the process and measurement models. The success of how well an estimator will perform thus lies within its foundation.

Niranjan and Cox (1994) for example used an autoregressive (AR) model closely related to speech signals for an internal model (*Level 1*). They then combined the results of various Kalman filters for each particular frequency channel using a pseudo-Bayesian algorithm (GPB1)(Bar-Shalom and Li, 1993) to track formant frequencies (*Level 2 & 3*). Hanekom and Krüger (2000) and Gray et al. (1994) used white process noise and coloured measurement (*Level 1 & 2*) noise processes to estimate the ISI. The shortcoming of this approach is that the process and measurement noise models cannot account for the entire process from source to spike train generation. Snyder (1975) on the other hand rather than tracking ISI used a Kalman-Bucy filter to track doubly stochastic Poisson point processes coding an intensity process (*Level 1 & 2*). Once again the spike trains were assumed to be Poisson distributed, something that is not necessarily true as will be discussed in *Chapter 3*.

Utilising a forward-predictive and sensory feedback model forces one to find a more pre-

cise model for the auditory system, by modelling it as close as possible to the physiology. Finding a biologically plausible system based on the most well understood parts of the auditory system, the outer, middle and inner ear including the neural coding just before the cochlear nucleus (CN), is thus the primary contribution that was made.

In short, the contributions made are:

1. The design of a forward-predictive and measurement model that can be linked to the auditory system from acoustic signal generation, right up to the neural processing functional unit shown in *Figure 1.11*,
2. to apply a Kalman filter to the forward-predictive and measurement model as was done for a motor system problem by Wolpert et al. (1995) and to examine the implications that such an implementation has and
3. finally to estimate with the help of an extended Kalman filter the power at a particular point within the frequency band, which forms part of a power map spanning across the entire audible frequency range that can be interpreted at higher levels in the CANS.

1.7 OVERVIEW OF DISSERTATION

The dissertation is logically divided into four main parts. The first part examines the physiology of the auditory system ranging from the outer ear right through to the inner ear and CANS.

In the second part the knowledge gained from the first part is applied and mathematical models are found describing the transition of the sound as it passes from the outer ear right through to the CANS. The group of models form the auditory system model (see *Figure 1.9*) that is used to generate the simulated sensory data.

In the third part of the dissertation the estimator equivalent model, as is depicted in *Figure 1.9*b, is designed by finding physiologically plausible forward-predictive and sensory-feedback models.

In the forth and final part of the dissertation the estimator model with forward-predictive model and sensory-feedback model is used to predict the power spectral map given noisy

measurements taken from the simulated auditory system, which is driven by a source. The entire system is simulated and the results and the implications thereof are discussed.

These four parts were subdivided amongst the following chapters in such a way, that a logical transition from gaining background knowledge on the subjected matter to designing and implementing the complete system can be made. A quick outline of the dissertation is given below.

Chapter 2 identifies the logical processing units found in the auditory system modelling the outer, middle, inner ear, in particularly the inner and outer haircells, and parts of the CANS. The physiological structure of each part of the auditory system is examined in order to characterise it and to identify constraints. In this way, more knowledge is gained about the biological implications that the auditory system has on modelling and simulations thereof.

In **Chapter 3**, the functionality of the inner haircells in terms of neural coding of information is discussed. Seeing that a lot of controversy exists around this part of the CANS, an in-depth look will be taken at the literature published thus far and their resulting implications.

Given the information about the physiology of the auditory system, mathematical models for each of the functional units identified in *Chapter 2* can now be derived and are found in **Chapter 4**. The model of the inner haircell was not included in this chapter as a clear distinction between physical and neurochemical transmission was made. This chapter also focuses on the interface between the surrounding area and the auditory system, such as a speaker generating acoustic signals, which are received by the auditory system and also defines what type of source signals are to be expected.

Chapter 5 investigates the transmission medium of the system, specifically the neural pathways found in the CANS. This is the only logical functional unit within the system that is closely related to the physiology rather than just the functionality of the auditory system. The channel, being the limiting factor in any communication system, is defined in terms of the way by which information is communicated from the one end to the other and the nature of the noise present. It also discusses how information is converted from a mechanical signal to APs by means of a leaky integrate-and-fire model (LIF) and how the CANS might interpret this information. In particular, it is examined how the signal can be decoded by means of optimal estimation and postsynaptic current (PSC) filtering.

In **Chapter 6** a state-space equivalent model for the auditory system (*Chapter 4* to *Chapter 5*) is defined and an extended Kalman filter (EKF) designed. A step-by-step design procedure for process and measurement noise models is given where a stochastic, zero-mean equivalent model representing the source signal, defined in *Chapter 4*, is also designed as part of the process noise model.

In **Chapter 7** the forward-predictive model from *Chapter 4* and the estimator equivalent model from *Chapter 6* are combined. The system is then evaluated in terms of its performance under noisy conditions found on the neural transmission channels. In particular it is investigated how spike-time jitter and the loss of spikes influence the performance of the EKF. The results and the general implications of the analysis-by-synthesis approach and the Kalman filter model are discussed. A link to the physiology is provided by comparing the just-noticeable difference in intensity discrimination of the model to psychoacoustic data in the literature.

In **Chapter 8** the dissertation is concluded by addressing the hypotheses and research question in context with the results obtained. The implications that these results have on the biology are also discussed. In addition, the strengths and weaknesses of the auditory system model and accompanying estimator model are stated and finally possible future work is presented.

CHAPTER 2

FUNDAMENTALS OF HEARING

2.1 AIM OF THIS CHAPTER

In this chapter the reader is provided with the necessary biological background of the auditory system and the functions of the outer, middle and inner ear. The transduction of mechanical information to neural stimuli is investigated and the knowledge gained here is later applied in *Chapter 4* to design a model for the auditory process. Although a short introduction is given to the type of spike patterns found in the CANS, particularly *rate-level coding* and *phase-locking*, the detail of what *information* is being transmitted is left for *Chapter 3*.

2.2 INTRODUCTION

Even though the analysis-by-synthesis approach was taken to derive a generic model to decompose, (en)code, transmit and then recover the original signal, each functional unit was derived by keeping in mind the physiology and functionality of the auditory system. It is thus imperative to define how each functional unit relates in certain ways to the biology. In this chapter a general overview of the entire auditory system is given, whereas the particulars of each functional unit within the system derived will be looked at in more detail in the successive chapters.

2.3 THE OUTER, MIDDLE, INNER EAR AND CENTRAL AUDITORY NERVOUS SYSTEM

The structure of the outer, middle and inner ear is depicted in *Figure 2.1*.

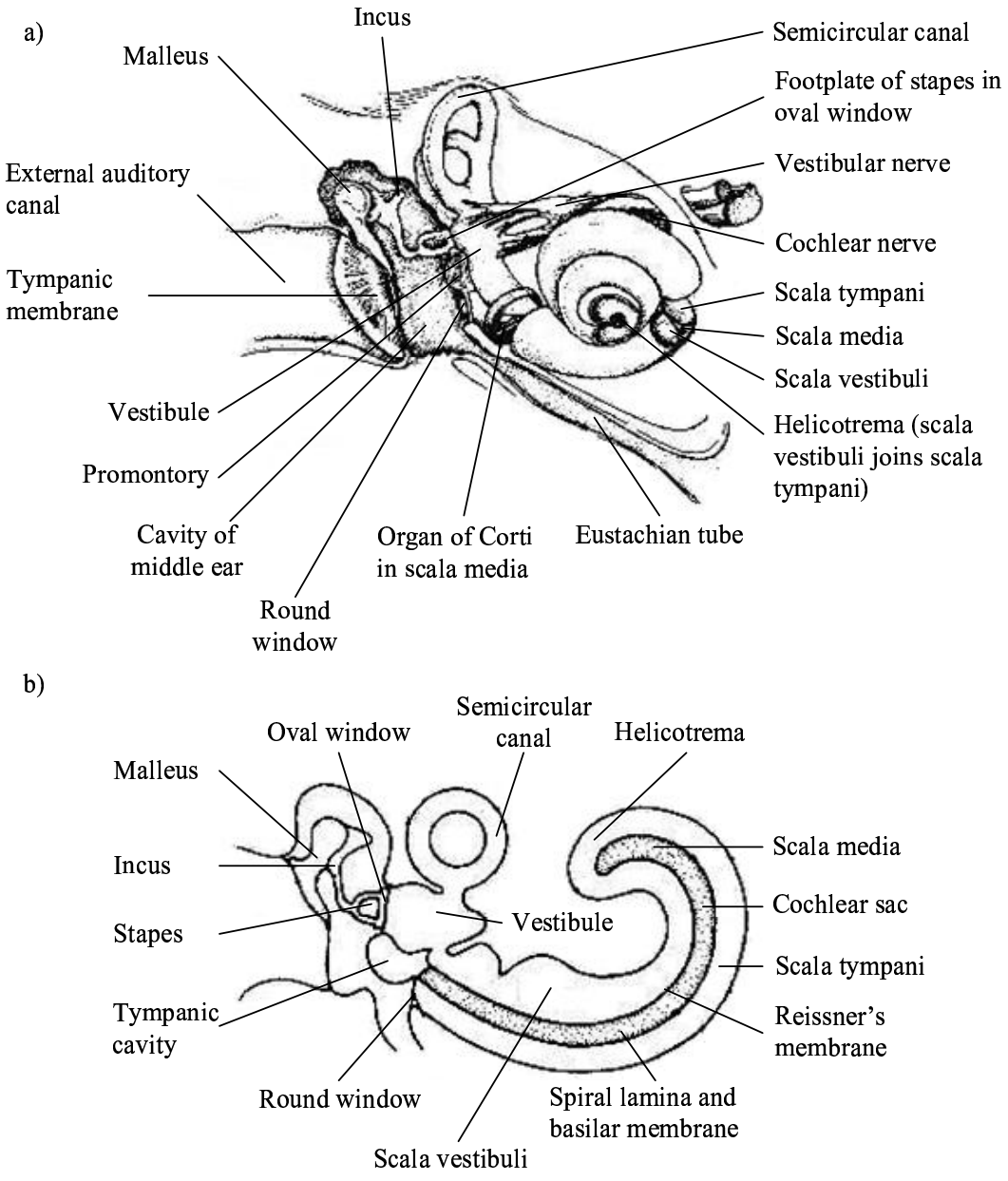

Figure 2.1: a) Main components of the inner ear in relation to the other structures of the ear. b) Schematic diagram of the middle ear and partially uncoiled cochlea, showing the relationship of the various scalae. Reprinted with permission by Yost (2000).

The auditory system can be divided into four sections: The outer ear, middle ear, inner ear and central auditory nervous system. When sound impinges on the *pinna* it gets funnelled by

the external auditory channel to the *tympanic membrane*. The tympanic membrane denotes anatomically the boundary between outer ear and middle ear. It is attached to the three middle ear bones collectively called the *ossicles*. The tympanic membrane is connected to the *malleus* via the handle, which in turn connects to the *incus*, which is connected to the *stapes*. The inner ear again can be divided into three parts: The *semicircular canals*, the *vestibule* and the *cochlea*.

The function of the middle ear is amongst others to pass the sound pressure acting on the tympanic membrane to the inner ear's *oval window*. This in turn physically excites the fluids within the inner ear, which results in a oscillatory movement of the *basilar membrane*. The movement of the basilar membrane causes the *cilia* of the haircells to vibrate, which translate this movement into neuronal action potentials. The coded sound signal is then transmitted from the cochlea via afferent bundles of neurons to the cochlear nucleus and beyond.

2.3.1 The Basilar Membrane and Resultant Neural Coding

2.3.1.1 Basilar membrane

To understand the principles behind the mechanical to neuronal action potentials conversion, a closer look at the mechanical response of the basilar membrane has to be taken.

The basilar membrane response is a travelling wave that is induced through the movement of the fluids within the inner ear responding to the pressure transmitted via the middle ear. The motion of the membrane travels towards the *helicotrema*.

The envelopes of the travelling waves for different stimulatory frequencies and their corresponding phase response are shown in *Figure 2.2*.

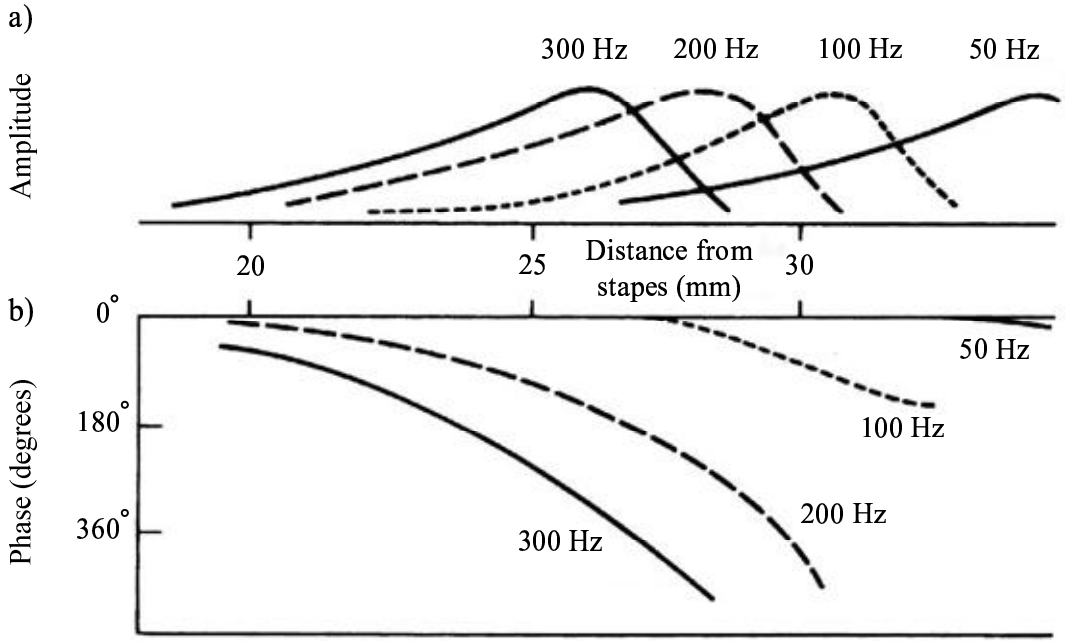

Figure 2.2: a) Envelopes of travelling waves of four different frequencies b) Phase shift in degrees between the motion of the stapes and a point on the basilar membrane. Reprinted with permission by Yost (2000).

While low excitation frequency stimulate the basilar membrane closest to the apex, high frequencies do so, closest to the base of the basilar membrane (Greenberg, 1997).

Similar amplitude and phase responses for the basilar membrane can be found in Flanagan (1972). It is of great interest to note that the amplitude response for a single frequency results in a Gaussian-like distributed curve around the stimulus frequency. This behaviour of course can be directly attributed to the nature of travelling waves propagating via the basilar membrane as will be discussed in *Chapter 4*. This implies that a single frequency excites multiple parts of the basilar membrane and thus groups of haircells, which fire with different group-delays.

2.3.1.2 Haircells and auditory nerves

There are approximately 12 000 outer and 3 000 inner haircells where the primary function of the outer haircells is to sharpen the frequency selectivity while the inner haircells carry

the neurally coded basilar membrane movement to the cochlea nucleus. When the basilar membrane vibrates, the *tectorial membrane* situated on top of the *organ of Corti* moves. Due to different hinging points, shearing forces between them are created. The *stereocilia* of the outer haircells, which are situated in the *organ of Corti* are directly connected to the tectorial membrane and are stimulated via bending. The result is an active response of the outer haircells to the stimulus, which consequently results in the expansion or contraction of the stereocilia also known as *motility*. The stereocilia of the inner haircells are probably not attached to the tectorial membrane and fluids trapped between the stereocilia and tectorial membrane probably cause inner haircell shearing (Yost, 2000). The action potentials generated by this particular form of bending are then transmitted via the cochlear nerve to the *cochlear nucleus*. Each cell needs a certain time to recover between firings. The neural response of auditory nerves firing to small movements of the basilar membrane show both intensity coding as well as phase-locking.

The knowledge of how the inner haircells code basilar membrane movement can be used to improve cochlear implants. Several research teams have tried to model such auditory neurons resulting in models such as the Brachman-Payton, Cooke and Meddis model (Meyer-Bäse and Scheich, 1997).

The type of neural coding scheme possibly used by the central auditory nervous system has been extensively analysed and potential models have been derived (Flanagan, 1972; Rieke et al., 1997; Maass and Bishop, 1999; Lestienne, 2001). A look is taken at two possible types of code, the *rate-level* code and the *phase-locking* code found in the auditory system. In *Section 3* a more extensive look will be taken at the neuronal code in general.

2.4 RATE-LEVEL FUNCTIONS IN THE AUDITORY SYSTEM

When increasing the level of the acoustic stimulus and measuring changes in the spike rate of a single neuron it is found that an increase in the intensity results in an increase in neuronal firing. The rate-level function is obtained by measuring the dependence of the spike rate on the intensity of the stimulus *averaged* over a period in time.

Several such rate-level functions are shown in *Figure 2.3*.

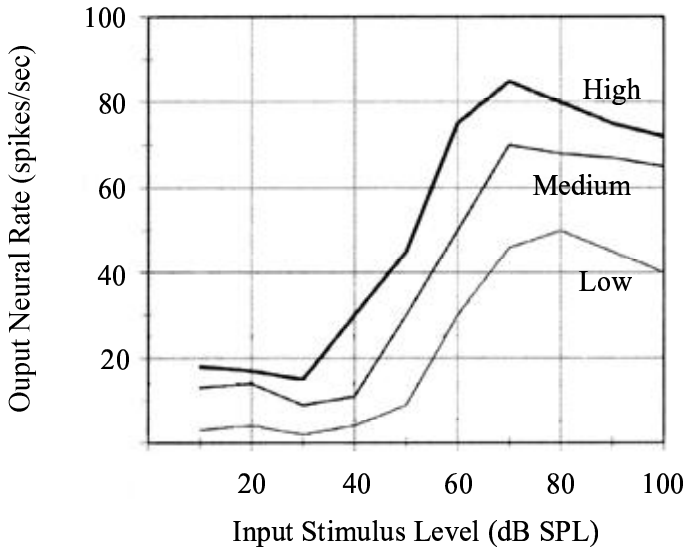

Figure 2.3: Intensity or rate-level functions for the three neurons: High, medium and low spontaneous rate fibers. The output is plotted in spikes per second as a function of the input level in dB sound pressure level (SPL). Reprinted with permission by Yost (2000).

The neural output spike rates for high, medium and low spontaneous rate fibers increase approximately linearly over a 30 to 70 dB sound pressure level (SPL), which is also known as the *neuronal dynamic range*. The rate-level function is never zero due to *spontaneous activity* by the neurons, resulting in a threshold spike rate or *neuron threshold*, which is present even if no stimulus is applied. Rate-level coding appears to be predominant at high frequencies in the auditory system taking over from the phase-lock coding above roughly 1 kHz.

2.5 PHASE-LOCKING

While the rate-level function is obtained by averaging over a time window and thus classified a *rate* code, *temporal* codes convey information by means of spike positions in time. When a sinusoidal stimulus is applied to the auditory periphery spike clustering is observed around a specific phase when phase-locking occurs. From the phase the fundamental frequency of that particular signal can be determined. At high frequencies clusters can also occur at submultiples of the spike rate due to a maximum possible spike rate of 500-1000 spikes per second per neuron. The degree of phase-locking is determined by how close the fundamental frequency corresponds to the characteristic frequency (CF) of that particular neuron.

Phase-lock coding is the principal coding mechanism found at low frequencies in the auditory system. The degree of locking seems to decrease for increasing frequencies and appears to disappear at 1 kHz (10% phase-locking) where the rate code takes over as the chief coding mechanism.

2.5.1 Current Computer Models

Computer models that simulate the function of the auditory system such as the one by Patterson, Allerhand and Giguere (1995) illustrate the neuronal activity graphically. *Figure 2.4* shows the response to the stimulus (vowel */e/*).

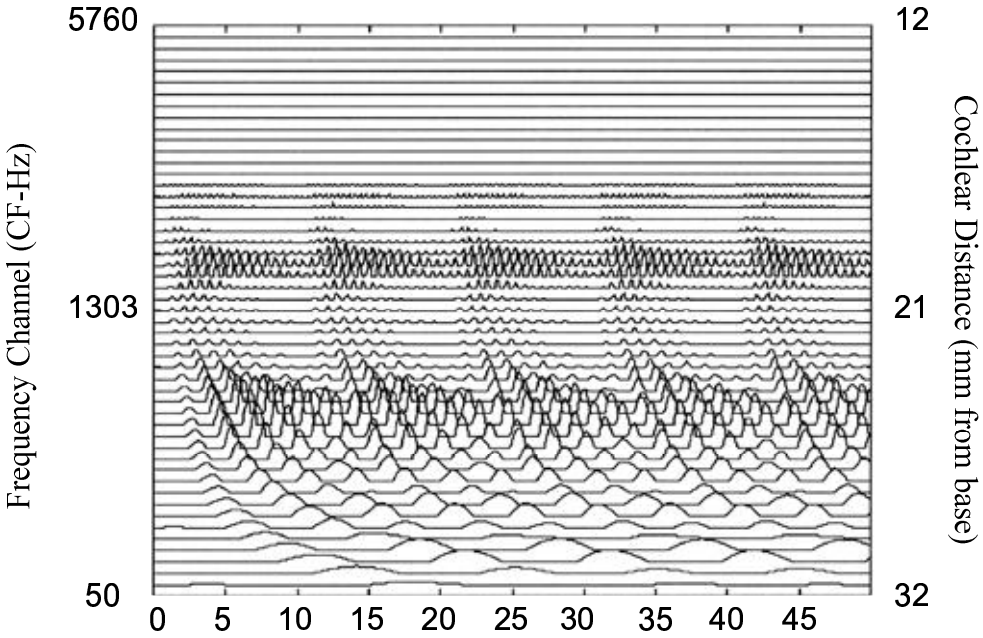

Figure 2.4: The output of a computational model of the auditory periphery. Each line is a simulation of a PST (Poststimulus Time) histogram of a tuned auditory nerve fiber. The right-hand y-axis represents the distance in millimetres from the stapes. Printed with permission by Yost (2000).

The frequency band is divided into a tonotopic fashion where each increment represents a CF along the distance of the cochlear.

2.6 CONCLUSION

This chapter examined the physiology of the auditory system and how each part of this system, from the outer ear right through to the inner ear and its outer and inner haircells, processes the external acoustic signal. This facilitates better understanding of the signal processing that takes place within the auditory system and allows for modelling by means of existing signal processing tools. Although not much information about the way *information* is being coded by the haircells was given, a better understanding of the fundamentals of neural coding in the cochlea in terms of *rate-level* coding and *phase-locking* was gained. In the next chapter a more critical look will be taken at neural coding in general.

CHAPTER 3

UNDERSTANDING THE NEURAL CODE

3.1 AIM OF THIS CHAPTER

Chapter 2 discussed how the inner haircells respond to vibrations of the basilar membrane. Two main trends were identified, *rate-level* coding and *phase-locking*. These are measurable responses of the inner haircells to basilar membrane stimulation. New research however shows that neural coding is much more complicated than was thought.

The aim of this chapter is to provide the reader with a summary as far as neural coding in the CNS is concerned by examining the advances that were made in the field of neuroscience over the last few decades. It is also examined how models for neural coding have changed and how different models explain certain aspects of neural coding found in the CNS but not others. It is thus by no means clear how the auditory system encodes information at the inner haircells but by means of a thorough investigation, certain assumptions can be made that are quite plausible. The relevance of these assumptions for this particular problem is discussed, a particular coding strategy chosen and then applied in *Chapter 5* where neural coding by means of a particular coder is evaluated.

3.2 INTRODUCTION

By means of neural coding the cochlear can convey the message it has received from the outside world to the CANS. *Chapter 2* discussed that a mechanical to neural action potential transduction takes place in the inner ear where the basilar membrane response is converted to a pattern of spikes or action potentials comprehendible to the CANS. In *Figure 1.5*, the spike generation and decoding process forms part of the Kalman filter model seeing that any measurements taken by the CANS are derived from action potentials. Measurements in the form of action potentials are however awkward to work with and converting them into a more

comprehendible format (for us) is desirable. A model representing the coding and decoding process thus needs to be found. In this chapter spike coding/decoding is discussed in terms of many possible models that have been documented by others.

Over the last century an enormous amount of information has been gathered and impressive progress has been made in the field of neuroscience. Man's inquisitiveness into the way that sensory neurons communicate information and the interpretation of this information has led to the inception of a vast number of models mimicking all stages of the neuronal communications processes of encoding, transmission and decoding.

In order to achieve an understanding of this code the underlying characteristics, given certain biological constraints, have to be studied. A look will be taken at the most important models and methods that have accumulated over the years. In *Chapter 2* the auditory system, the mechanism of mechanical to neurological transduction and possible codes found in the auditory system were briefly discussed. The possible coding scheme, implemented by the physiology has been the focus of many years of research with the most notable publications by the first pioneers such as Pitts and McCulloch (1947), MacKay and McCulloch (1952), Hodgkin and Huxley (1952) to more recent publications by Rieke et al. (1997), Maass and Bishop (1999), Martignon, Deco and Laskey (2000), Lestienne (2001), Eggermont (2001), Eliasmith and Anderson (2003).

3.2.1 Neural Coding Background

Early studies of spike trains showed that under repeated trials (Yost, 2000; Deutsch and Deutsch, 1993; Lestienne, 2001) the average spike count over an appropriate window was directly proportional to the intensity of an input stimulus. These studies also suggested that the emission of sensory fibers was not completely deterministic and that a component of stochasticity was present. Experiments showed that this random behaviour could be described by a Poisson point process and that the underlying information was a spike rate, corrupted by noise (Teich and Khannan, 1985; O'Neil, Lin and Ma, 1986; Johnson, 1996). It has however come to light over the last decades that a rate code might be insufficient to describe the coding of information in the CNS (Lestienne, 2001; Rieke et al., 1997; Martignon et al., 2000). It has been evident from more recent studies that the relatively large averaging windows required to determine the average spike rate are biologically implausible. Reaction times such

as the compensating flight torque by the fly have been measured in the order of milliseconds, where cells firing 100 spikes per second can generate only a few spikes (Rieke et al., 1997). Further experiments involving the bat auditory system also underlined the importance of small numbers of spikes. It has further been shown by Frisina (2001) and Eggermont (2001) that both rate and temporal codes are found in the CNS implying that at different levels different codes could be present, each fulfilling a particular role.

The relevance of a neural code and how to chose a particular one for design purpose is however not a simple task due to the ambiguity that exists in the CNS about what type of neural coding is utilised. In *Section 3.8* this problem is addressed more closely, but before such a decision can be made a closer look at *rate* codes and *temporal* codes has to be taken.

3.2.2 Coding versus Encoding

While common neuroscience literature use the terms *coding* and *encoding* synonymously, they are in fact implying two different processes (Lestienne, 2001). While temporal *coding* for example is characterised by a one-to-one correspondence between the time of occurrence of a sensory event and the time of occurrence of an action potential, temporal *encoding* of a signal corresponds to situations in which information of a dynamic signal is encoded in a temporal pattern of action potentials. Throughout this document it will be assumed that an input stimuli can only be *coded*. This however does not imply that encoding schemes are not implemented by the CNS.

3.3 RATE CODE

A rate code conveys one parameter at a time, the *average rate*, which is obtained by determining the average number of spikes in a particular window size, where the window size can be variable (typically 100 ms for a maximum spike rate of 300-500 spikes per second) (Rieke et al., 1997). An interesting statistically characteristic of a rate code is that it becomes a time code if the window size shrinks to zero. A rate code can also be described stochastically by a probability of spike occurrences.

With a rate code a tradeoff exists between reaction time of a code and its robustness to noise. This tradeoff is determined by the size of the window over which the average number

of spikes is determined. For example if the window size is made large, the code is slow to react to fast stimuli but provides a robust estimate in the presence of noise. On the other hand, if the window size is made small, the spike rate becomes more prone to interference from noise but reacts faster to fast changing stimuli.

A clear disadvantage of the rate code is that it cannot code more than one variable at a time, making it a one-dimensional code with a low transmission rate. This is not the case as far as temporal codes are concerned as is evident from the next section. Another interesting fact about the rate code is that under repeated trials of the same input stimulus different temporal structures were produced that can be attributed either to noise or to higher-order underlying statistics (Eliasmith and Anderson, 2003).

As was mentioned in *Subsection 3.2.1* rate coding cannot explain the fast reaction times that are possible in the bat auditory system, where large averaging windows would introduce unrealistic delays. This however does not mean that pure rate codes do not exist since they play a vital role in the auditory system after the CN and the central auditory midbrain and the cortex (Frisina, 2001). Thus a rate code can explain some parts of neural coding but not others and seems to be a particular type of code used selectively by the CNS.

3.4 TEMPORAL CODES

A temporal code, unlike a rate code, assumes that information is carried in the *precise timing* of spike positions. There are two major types of temporal codes (Eliasmith and Anderson, 2003):

1. The time code or instantaneous rate code depends on the interspike interval (ISI) and
2. the second type depends on the placement of a spike relative to the onset of a stimuli.

While the former is based on a group of spikes like *phase-locking* and *real-zero crossing* (RZC) (Kumaresan and Wang, 2001) codes, the latter depends entirely on just one single spike and its position relative to the onset of the stimulus. Codes depending on ISI are typically used to code more than one parameter at a time (see *Subsection 3.4.1* and *Subsection 3.4.2*) where single spikes are used as *classifiers* in spiking neural nets (Maass and Bishop, 1999).

Temporal codes have been compared to rate codes on various occasions (Rieke et al., 1997; Rullen and Thorpe, 2001) using an information transmission approach. The rate code was compared to the mean interspike interval code and it was found that the rate code is far from optimal for fast information transmission whereas temporal structure of spike trains can be used effectively to maximise information transfer rates when each cell fires only once. Temporal codes however are far less robust in the presence of noise.

In *Section 3.3* it was mentioned that a rate code could not explain certain parts of neural coding such as fast reaction times given only a few spikes. The same ambiguity exists with temporal codes. It has been shown that precise timing is not always required for successful transmissions (Rieke et al., 1997). Ones again it seems that the CNS uses different codes for different tasks.

3.4.1 Phase-Lock Coding

Phase-locking in the auditory system was discussed in *Section 2.5*. Being one of the two forms of temporal codes that have been defined thus far provides a mean of encoding both amplitude and frequency information rather than just amplitude information as the rate code does. It has been suggested by Rieke et al. (1997) that the instantaneous rate of firing can be linked as following to amplitude (A), frequency (ω) and phase (ϕ) information:

$$r(t) = r_0 + GA\sin(\omega t + \phi). \tag{3.1}$$

The spontaneous spike rate r_0 can be assumed to be large enough so that the spike rate at any given time will never be smaller than zero. The desired increase in spike rate relative to the input amplitude can be scaled by any desirable constant G. Low spike rates at the troughs of the sinusoid are however a matter of concern in that low spike rates are more susceptible to noise especially if only one channel carries the information. The physiology of the auditory system avoids such problems at low frequencies by employing populations of neurons.

It is interesting to note that the instantaneous spike rate was used to encode temporal information. This suggests that it is not the *spike rate* that makes a code a rate code, rather it is the way the spike rate is determined. If the instantaneous spike rate was however averaged over a window larger than the period of $r(t)$, no phase-locking would have been evident even

though it was present.

3.4.2 Representing Signals by using only Timing Information

A recent publication by Kumaresan and Wang (2001) showed that a signals envelope and phase could be represented by timing information alone. While only real-zero (RZ) signals can uniquely be represented by their zero-crossings it was shown that any band-pass filtered signal could be mapped to such signals. The result is that any signal can be spectrally decomposed by a filter bank and then coded by a real-zero crossing (RZC) algorithm, resulting in a spike train or in this case known as the converted zero crossings (CoZeCs), representing the zero-crossings of the mapped real-zero signals. A synthesis filter bank then reconstructs the spike trains into an estimate of the original signal. It was also shown that this particular encoding scheme performs well if no noise is present on the spike trains, but did not pursue the matter further than that. A disadvantage of this method however is that two pairs of CoZeCs need to be transmitted, requiring four channels per spectral block. For a system that guarantees no reliability in the presence of noise, it requires a particularly high number of neurons.

3.5 POPULATION CODES

Recent advances in multi-electrode recording have brought us closer to the understanding of population codes. They have been observed and studied from amongst others in the retina (Nirenberg and Latham, 1998) to the hippocampal place cells in rats (Zhang, Ginzburg, McNaughton and Sejnowski, 1998). In these cases a neuron occupies a "place" where spatial overlapping results in information sharing, which has been shown to increase robustness of transmission. Population codes convey no information about the exact timing of spikes but weight the spike rate by a means of a tuning curve. Tuning curves specify the average response of a cell to a certain input as a function of the feature values of the cell. Population of neurons can be shown to implement optimal estimators such as maximum likelihood estimators (MLE) (Zhang et al., 1998).

The response variability of the neurons within a population have a major impact on the encoding capacity of the group. It has been shown by Pouget, Zhang, Deneve and Latham (1998), Brunel and Nadal (1998) and Pouget, Deneve, Ducom and Latham (1999) that narrower tuning curves increase the Fisher information if the noise distribution is constant over

the width of the tuning curve. The Fisher information determines how well an estimate of a parameter can be found from an observation with a given probability distribution function. Wilke and Eurich (2001) showed that representational accuracy of populations with inhomogeneous tuning properties such as tuning curve width and fragmentation into specialised subpopulations were superior to identically and radially symmetric tuning curves. Zemel, Dayan and Pouget (1998) argued that a population activity can convey higher-order statistics and not just a single value.

While population coding automatically assumes interaction due to cross-correlations between neurons within a group, studies by Nirenberg and Latham (1998) showed that linear reconstruction methods, which treat the cells as independent coders, were as effective as a neural network that can utilise correlations amongst different cells when reconstructing the stimulus. The same observation was made earlier by Warland, Reinagel and Meister (1997) where experiments suggested that correlated activity plays a minimal role in encoding visual stimuli. That however does not imply that adjacent neurons within a group are completely independent since a population of neurons has been shown to convey information reliably within a short period of time (Panzeri and Schultz, 2001), something that an isolated rate-coding neuron is unable to do.

It is thus apparent that the way that neurons, as individuals or a group, code and reconstruct information differently depending on the requirement of a particular system.

3.6 SPATIOTEMPORAL CODING

Unlike population codes, spatiotemporal grouped neurons utilise the *timing* information of the spikes found on each nerve to make a decision rather than the spike rate.

Martignon et al. (2000) found that higher spatiotemporal order statistics within a group (more than 2 neurons) of neurons was plausible if log-linear models were used. The results were applied to detect synchronisation in experimental data, in particular the recordings for the frontal cortex of Rhesus monkeys. The experiments confirmed the existence of synchronisation and thus higher-order temporal patterns, which in turn confirms that *encoding* can take place in the CNS.

Deco and Schürmann (1999), from the information point of view designed a model that represents a group of neurons that need to classify a particular input, given spike-time information. However, instead of finding an optimal estimator, as suggested in *Section 5.4.2*, they developed a neural net that trains the weights between the neurons and the axonal transmission delay related to a particular weight. In effect synchronous clusters are formed that fire when a particular input stimuli is present.

Masuda and Aihara (2002) did similar experiments using integrate-and-fire (IF), leaky integrate-and-fire (LIF) and FitzHugh-Nagumo (FHN) neurons and a multi-layer neural network where a group of sensory neurons drive a cortical neuron. The spike trains from the different sensory neurons were superimposed and shown to interfere with each other, but not to such an extent that it became detrimental. The same spike train superimposition method was implemented by Hanekom and Krüger (2000). Masuda and Aihara (2002) used the ISI to reconstruct the original continuous signal by means of a filter. They showed that the coding scheme was quick in encoding changes in an input signal with little energy required. Similar rate codes would be slower in reacting to fast changing stimuli and require higher spike rates and thus more energy to code.

3.7 ADAPTIVE NEURAL CODING

In order to code efficiently, with regards to information theory, the coding strategy has to match the statistics of the signal. This would require an adaptive code instead of a static one as was assumed in the coding strategies discussed previously. The relevance of adaptive neural coding can be understood better by considering the following reasoning. In the case of a H1 cell of the blowfly the dynamic range spans less than two orders of magnitude where the ambient light level varies over more than nine (Smirnakis, Berry, Warland, Bialek and Meister, 1997). The neural circuits are faced with the trade-off between encoding the full range of their inputs and resolving gradations among those inputs. It was shown by Smirnakis et al. (1997) and later by Brenner, Bialek and de Ruyter van Steveninck (2000) and Fairhall, Lewen, Bialek and de Ruyter van Steveninck (2001) that light adaptation takes place where the sensitivity of the retina becomes proportionally less as the intensity of the light increases. It was further shown that not only changes in the mean light intensity were provided for but also changes in the variance. Adaption to these statistics occurred over a wide range of time scales, from a few tens of milliseconds up to a minute. By adapting to the statistics of the environment

the information transmission rate is also maximised as was observed by Brenner et al. (2000).

Adaptive neural coding however introduces ambiguities since the meaning of spikes depend on the context. Thus the additional requirement imposed on adaptive neural coding is context coding, which can be done by coding the context separately or by using a interval distribution method devised by Fairhall et al. (2001). It was shown that by collecting interspike intervals and finding their interval probability density, a distinct "fingerprint" for fixed zero mean and variable variance Gaussian random variables could be identified.

Adaptive neural coding illustrates that the neural coding system found in various parts of the CNS is indeed an efficient one, maximising the transmission rate for different input signal statistics.

3.8 CHOOSING A CODE

As is quite evident from the preceding section, choosing the "best" or "correct" code, if there is such a thing, is not a trivial task. In fact it has been argued by Frisina (2001) that in the auditory system there exist various levels of coding between the cochlear hair cells from where spikes originate up to the auditory periphery and cochlear nucleus (CN) and also the auditory midbrain and cortex. It seems that synchronisation is important up to the auditory periphery, whereas rate coding is more important for the auditory midbrain. Frisina (2001) also stated that different types of neurons, low, medium and high spontaneous rate fibres encoded high, medium and low sound intensities respectively, which agrees with *Chapter 2*.

In principle, for simulation purposes, picking the rate code as the fundamental code is an adequate choice as long as all constraints of the particular choice are defined and taken into consideration. The rate code is one of the most widely applied codes in neural coding simulations and forms part of recent publications such as Eliasmith and Anderson (2003). While a rate code is robust in a noisy environment if the averaging window is chosen to be large, it cannot code more than one variable at a time and coding both amplitude and phase for example would require two dedicated neurons. A temporal code such as phase-locking however can code both amplitude and phase and thus increase the amount of information that is being transmitted by one neuron. Robustness plays an important role in this problem since the power spectral map needs to be estimated as best as possible. However, as was mentioned

in *Section 3.6* exploiting cross-correlation between neurons in the decoding process is no more effective than treating them as independent neurons as far as linear reconstruction methods are concerned. Thus in the case of linear decoders no advantage is gained by using more complex cross-correlated neuron networks. For this reason it seems natural to assume that no cross-correlation exists between different neurons.

As was discussed in *Section 3.7*, no one particular rate or temporal code can optimally code a signal with changing input statistics. This is an important aspect of neural coding since the auditory system can effectively work with a wide variety of acoustic signals ranging from pure tones to complex speech patterns. It is thus evident that the spike coding/decoding functional unit in the CANS should be a *dynamic* rather than a static one, adapting to new input signal statistics as they become available. This would however require another observer system, keeping track of the statistics of the signal being estimated. Thus in order to avoid overcomplicating the problem a static spike coding/decoding model was chosen.

From the discussion above it is evident that in order to pick a neural code the requirements of the coding system based on the advantages and disadvantages of a particular code, needs to be defined. Since the definitions of the different neural codes however already lead to confusion, it is refrained from picking a neural code as such. Instead, in *Chapter 5* a particular *coding mechanism* is chosen based on physiological information that has been discovered and documented. The implications of such a decision leads to an interesting discovery about neural codes.

3.9 CONCLUSION

In this chapter the different types of neural code were introduced and discussed. Each code was examined in terms of how information is transmitted by means of trains of action potentials both on a single neuron and in groups. The advantages and disadvantages of each code were mentioned and compared. Since the boundary between a rate code and a temporal code is not clearly definable and disputes continue, no particular code is chosen here. Alternatively, biologically plausible neuron models are investigated and complementing linear estimators defined. On those grounds and based on the information found in this chapter it will be briefly discussed, which theoretical definition of a neural code fits this specific implementation the best. The motivation for such an approach is to demonstrate to the reader the inconsistencies

within the *definitions* of a neural code.

CHAPTER 4

SIGNAL DEFINITION AND DECOMPOSITION

4.1 AIM OF THIS CHAPTER

Now that information about the auditory system (*Chapter 2*) and neural coding process (*Chapter 3*) has been gathered, mathematical models can be derived. The aim of this chapter is to design and simulate a biologically plausible auditory system (representing the forward-predictive model of the auditory system in *Figure 1.5* in the logical form of *Figure 4.1*) from the outer ear right through to the inner ear, excluding the mechanism for mechanical to neural transduction.

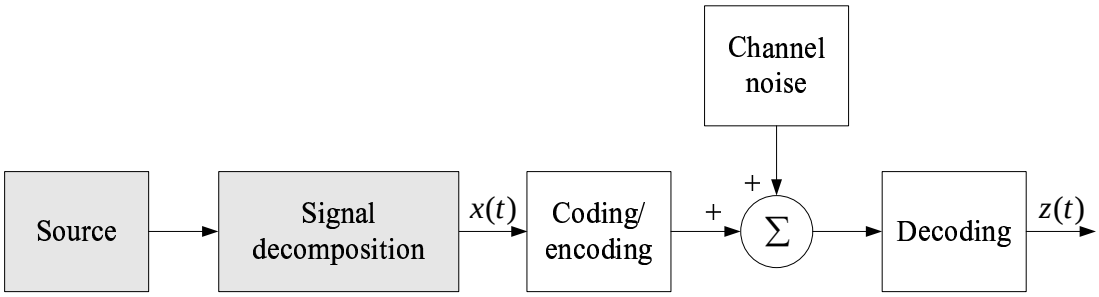

Figure 4.1: The shaded functional units, *source* and *signal decomposition* are designed and evaluated in this chapter.

The model for the inner haircells (represented by part of the spike coder/encoder functional unit in *Figure 4.1*) is designed in *Chapter 5*. This chapter also analyses the signal generation aspect, seeing that a good model for the acoustic signal is required in order to build an estimator in *Chapter 6*.

4.2 INTRODUCTION

In order to test the hypothesis that an internal model of the auditory system is present, an accurate signal model has to be found that describes the source driving the model. In the next few sections the requirements of the process model are investigated, which in this case is a simple pure-tone with time varying amplitude, frequency and phase.

4.3 INPUT SIGNAL DEFINITION

Speech represents the major complex acoustic stimulus for most humans. Speech waveforms can be viewed as a sum of sinusoids, which have temporal and spectral units. Frequencies created by the vibration of the vocal folds are called *voicing fundamental frequency* and frequencies at which the vocal tract resonates are called *formant frequencies*. A *phoneme* is the basic auditory unit of speech allowing for the identification of the difference between the words had, head, heed and hid. Most phonemes differ in the various frequencies of the various formants. While for most vowel sounds the formant frequencies stay constant over time, in the case of most other phonemes various formant frequencies do change. These changes however occur relatively slow in time and formant frequencies can stay constant for between approximately 10 ms to 100 ms (Cole, 1980; Yost, 2000). Another interesting phenomena is that no more than 3 to 5 fundamental formant frequencies are required to correctly identify a phoneme (Cole, 1980). For example the vowel /e/ has formants located at approximately 500 Hz, 1.8 kHz and 2.6 kHz respectively (Yost, 2000).

Speech is a particular form of acoustic stimuli and can be divided into two distinct types (e.g., Haykin, 1994):

1. *Voiced sounds:* The source of excitation is pulse like and periodic where voiced sounds during vowels are characterised by quasi-periodicity, low-frequency content, and large amplitude.
2. *Unvoiced sounds:* The source of excitation is noise like, characterised by randomness, high-frequency content, and relatively low amplitude.

The next step is to simulate an acoustic signal based on a model of speech, which is defined below.

4.3.1 Input Signal Definition

The most complex signal the auditory system has to contend with, namely speech, can be represented by three to four formant frequencies spread across the frequency range in a particular manner. Speech can be subdivided into voiced and unvoiced sounds, as was already mentioned.

Creating a signal model to generate both voiced and unvoiced signals at the same time would be ideal, since it would represent the true nature of a speech signal. This would however require two processes, one quasi-periodic in the low-frequency range and one random in the high-frequency range. Common speech synthesisers switch between the two models depending on the type of speech signal being modelled (Haykin, 1994). To avoid such a complex model of the two stimuli types only the more complex voiced sounds are considered. The unvoiced sounds are easily simulated as shown by Moore (1972) and Haykin (1994) where a white-noise source drives a continuous-time first-order linear time-invariant system in the form of

$$X(s) = \frac{a}{s+a}. \tag{4.1}$$

While it is easier to model a random process stochastically, signals that are more deterministic and quasi-periodic by nature provide for a more challenging task in modelling. The signal model in this case is taken to be a close representation of the *envelope* of a formant (with harmonic structure) found in speech signals. Besides some slight alterations to the nature of a true formant, a strict set of conditions is defined. This simplifies the signal model without loosing the desired functionality, namely, a periodic signal changing randomly in amplitude, frequency and phase.

The input signal to the system is defined as being:

1. A single periodic sinusoid with varying amplitude ($\mathbf{r}$), frequency (ω) /2π and phase (ϕ) with time, which represents the envelope of a formant,
2. where the formant is constant over a specified time window before an instantaneous frequency "jump" occurs.
3. Amplitude changes occur simultaneously with the frequency and phase changes,

4. where both the amplitude and frequency are taken to be Gaussian random variables and the phase a uniformly distributed random variable.

5. The signals bandwidth is limited to 500 Hz.

The signal can be thought of as modulating a particular amplitude ($\mathbf{r}$) (measured in volts) with a carrier frequency $(\omega)/2\pi$ (measured in Hertz) by means of amplitude modulation (AM), as is discussed later in the section, where the magnitude and position (in the spectral band) stays constant for a period of time after which the *formant* frequency changes again. Although speech is normally band limited to 3.2 kHz as in the case of the analogue telephony, this speech signal was limited to 500 Hz for the following reason:

1. The speech signal model is practically more realisable due to a reduced bandwidth translating to a smaller number of samples.

In order to estimate the behaviour of the speech signal under noisy transmission conditions, no noise is specifically added to the signal itself. Although this in not a very realistic approach it serves a special purpose, namely to investigate how reliable the communications *within* the central auditory nervous system have to be in order to be able to estimate the stimulus to a certain degree. This is the common communications engineering approach to simulate the reliability of any transmission system, given a particular noisy or unreliable channel.

Since the characteristics of the speech signal model have now been defined, a stochastic model can be derived. The stochastic model forms an integral part of the internal model of the Kalman filter as was mentioned in *Section 1.4* and is derived below.

4.3.2 Stochastic Input Signal Analysis

The aim of analysing the source signal as a stochastic process is to facilitate understanding of the statistics of this speech signal. This paves the way to finding a shaping filter for the speech signal in *Chapter 6*.

Keeping this in mind the voiced sounds are defined as being a quasi-periodic random signal with a large amplitude and low frequency, stochastically defined as

$$\mathbf{x}(t) = \mathbf{r}\cos(\boldsymbol{\omega}t + \boldsymbol{\phi}), \tag{4.2}$$

where a bold typeface represent a continuous random variables, $\mathbf{r}$ represents the random amplitude, $\boldsymbol{\omega}$ the random frequency and $\boldsymbol{\phi}$ the random phase. If it can be assumed that the variables $\mathbf{r}$, $\boldsymbol{\omega}$ and $\boldsymbol{\phi}$ are statistically independent from each other with $\boldsymbol{\phi}$ being uniformly distributed over the interval $(-\pi, \pi)$, it can be shown that the autocorrelation function can be found by setting

$$\begin{aligned} \psi(t_1, t_2) &= E\left[\boldsymbol{x}(t_1)\boldsymbol{x}(t_2)\right], \\ &= E\left[\mathbf{r}\cos(\boldsymbol{\omega}t_1 + \boldsymbol{\phi})\mathbf{r}\cos(\boldsymbol{\omega}t_2 + \boldsymbol{\phi})\right], \end{aligned} \tag{4.3}$$

and given the trigonometric identity

$$\cos\alpha\sin\beta = \frac{1}{2}\left[\cos(\alpha - \beta) + \cos(\alpha + \beta)\right], \tag{4.4}$$

Equation (4.3) can then be solved as:

$$\psi(t_1, t_2) = E\left[r^2\right] E\left[\frac{1}{2}\left[\cos(\boldsymbol{\omega}(t_1 - t_2)) + \cos(\boldsymbol{\omega}(t_1 + t_2) + 2\boldsymbol{\phi})\right]\right]. \tag{4.5}$$

Using the trigonometric identity

$$\cos(\alpha \pm \beta) = \cos\alpha\cos\beta \mp \sin\alpha\sin\beta, \tag{4.6}$$

the fact that

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \cos 2\boldsymbol{\phi}\; d\boldsymbol{\phi} = 0, \tag{4.7}$$

and

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\sin 2\boldsymbol{\phi}\ d\boldsymbol{\phi} = 0, \tag{4.8}$$

the autocorrelation of the random process in *Equation (4.2)*, assuming only a uniformly distributed phase between $(-\pi, \pi)$, is then given by

$$\psi(t_1, t_2) = \frac{1}{2}E\left[\mathbf{r}^2\right] E\left[\cos\left[\boldsymbol{\omega}(t_1 - t_2)\right]\right]. \tag{4.9}$$

If the process in *Equation (4.2)* is wide sense stationary (WSS) then $\psi(t_1, t_2) = \psi(\tau)$.

The autocorrelation function of the random process (*Equation (4.2)*) can now be evaluated, given different distributions for the first and second expectation in *Equation (4.9)*. The autocorrelation function or the power spectral density (PSD) function uniquely describes any random process. In this case the autocorrelation function of $\mathbf{x}(t)$ in *Equation (4.2)* is of special interest. It describes how the *combination* of the random variables $\mathbf{r}$, $\boldsymbol{\omega}$ and $\boldsymbol{\phi}$ behaves as one single random variable. By determining $\mathbf{x}(t)$ in terms of the amplitude, frequency and phase an implicit solution is found for the mixing of random variables. Thus not three, but only one process needs to be simulated, simplifying the model considerably as will be evident in *Chapter 6*.

To investigate how the three random variables behave, appropriate distributions have to be found for each one of them. The derivations are limited to the Gaussian and uniform distributions resulting in a very useful outcome later. For the first term, only the Gaussian distribution will be evaluated while for the second term both the uniform and Gaussian distributions will be evaluated. In particular, if the amplitude $\mathbf{r}$ is a Gaussian random variable (RV) with $N(0, \sigma_{\mathbf{r}}^2)$, where $N(\boldsymbol{\mu}, \sigma^2)$ is defined as a Gaussian random variable with mean $\boldsymbol{\mu}$ and variance σ^2, then

$$\begin{aligned} E\left[\mathbf{r}^2\right] &= \int_{-\infty}^{\infty}\mathbf{r}^2\frac{1}{\sqrt{2\pi}\sigma_{\mathbf{r}}}\ e^{-\mathbf{r}^2/2\sigma_{\mathbf{r}}^2}\ d\mathbf{r}, \\ &= \frac{2}{\sqrt{2\pi}\sigma_{\mathbf{r}}}\int_{0}^{\infty}\mathbf{r}^2 e^{-\mathbf{r}^2/2\sigma_{\mathbf{r}}^2}\ d\mathbf{r}. \end{aligned} \tag{4.10}$$

Given that

$$\int_0^\infty x^2\ e^{-ax^2}\ dx = \frac{1}{4a}\sqrt{\frac{\pi}{a}} \qquad [\mathrm{a} > 0]\,, \tag{4.11}$$

it follows that

$$\begin{aligned} E\left[\mathbf{r}^2\right] &= \frac{2}{\sqrt{2\pi}\sigma_{\mathbf{r}}}\left[\frac{\sigma_{\mathbf{r}}^2}{2}\sqrt{2\sigma_{\mathbf{r}}^2\pi}\right], \\ &= \sigma_{\mathbf{r}}^2 \qquad \left[\sigma_{\mathbf{r}}^2 > 0\right]. \end{aligned} \tag{4.12}$$

If the second term in *Equation (4.9)* is a uniform RV with $E\left[\mathbf{x}\right] = 0$ and $E\left[\mathbf{x}^2\right] = f_c^2/3$ then

$$\begin{aligned} E\left[\cos(2\pi\mathbf{f}\tau)\right] &= \frac{1}{2f_c}\int_{-f_c}^{f_c}\cos(2\pi\mathbf{f}\tau)\ d\mathbf{f}, \\ &= \frac{1}{4\pi f_c\tau}\left[\sin(2\pi\mathbf{f}\tau)\right]\Big|_{-f_c}^{f_c}, \\ &= \frac{\sin(2\pi f_c\tau)}{2\pi f_c\tau}. \end{aligned} \tag{4.13}$$

If $\tau = 1/\mathrm{T}_s$, in other words, if the continuous RV is sampled at the *Nyquist rate* then $E\left[\cos(2\pi\mathbf{f}\tau)\right] = 0$ thus the samples will be statistically independent of each other. On the other hand if the second term in *Equation (4.9)* is a Gaussian RV with $N(0, \sigma_{\mathbf{f}}^2)$ then

$$E\left[\cos(2\pi\mathbf{f}\tau)\right] = \int_{-\infty}^{\infty}\cos(\boldsymbol{\omega}\tau)\frac{1}{\sqrt{2\pi}\sigma_{\mathbf{f}}}\ e^{-\boldsymbol{\omega}^2/2\sigma_{\mathbf{f}}^2}\ d\boldsymbol{\omega}, \tag{4.14}$$

and given that

$$\int_0^\infty \cos(bx)\ e^{-ax^2}\ dx = \frac{1}{2}\sqrt{\frac{\pi}{a}}\ e^{-b^2/4a} \qquad [\mathrm{a} > 0]\,, \tag{4.15}$$

Equation (4.14) reduces to

$$\begin{aligned} E\left[\cos(2\pi \mathbf{f}\tau)\right] &= \frac{2}{\sqrt{2\pi}\sigma_{\mathbf{f}}}\int_0^{\infty} \cos(\tau\boldsymbol{\omega})\ e^{-\boldsymbol{\omega}^2/2\sigma_{\mathbf{f}}^2}\ d\boldsymbol{\omega}, \\ &= e^{-(\tau\sigma_{\mathbf{f}})^2/2}. \end{aligned} \tag{4.16}$$

Thus for the uniformly distributed random amplitude, phase process and Gaussian distributed random frequency process the autocorrelation function is given by

$$\psi(\tau) = \frac{1}{2}\sigma_{\mathbf{r}}^2 \frac{sin(2\pi f_c \tau)}{2\pi f_c \tau}, \tag{4.17}$$

and for the Gaussian-distributed random amplitude, frequency process and uniformly distributed random phase process the autocorrelation function is given by

$$\psi(\tau) = \frac{1}{2}\sigma_{\mathbf{r}}^2\ e^{-(\tau\sigma_{\mathbf{f}})^2/2}. \tag{4.18}$$

From *Equation (4.17)* it follows that the autocorrelation function is a sinc function, decreasing with increasing τ. The zero-crossings of the autocorrelation function are at

$$\tau = f_c, \frac{1}{f_c}, \frac{2}{f_c}, \cdots \tag{4.19}$$

This has the implication that if *Equation (4.17)* is sampled at

$$T_s = \tau = f_c, \frac{1}{f_c}, \frac{2}{f_c}, \cdots \tag{4.20}$$

then the successive samples are uncorrelated and statistically independent.

Both stochastic models given by *Equation (4.17)* and *Equation (4.18)* are based on the assumption that the random variable $\mathbf{f}$ has a zero mean. In order to find a more general model with an arbitrary, non-zero mean μ for the random variable $\mathbf{f}$, the second term in *Equation (4.9)* is derived such that

$$E\left[\cos(2\pi \mathbf{f}\tau)\right] = \int_{-\infty}^{\infty} \cos(\boldsymbol{\omega}\tau)\frac{1}{\sqrt{2\pi}\sigma_{\mathbf{f}}}\ e^{-(\boldsymbol{\omega}\ -\ 2\pi\mu)^2/2\sigma_{\mathbf{f}}^2}\ d\boldsymbol{\omega}. \tag{4.21}$$

Performing variable substitution in *Equation (4.21)* by letting

$$\begin{aligned} u &= \boldsymbol{\omega} - 2\pi\mu \\ du &= d\boldsymbol{\omega}, \end{aligned} \tag{4.22}$$

Equation (4.21) now becomes

$$E\left[\cos(2\pi \mathbf{f}\tau)\right] = \frac{1}{\sqrt{2\pi}\sigma_{\mathbf{f}}}\int_{-\infty}^{\infty} \cos\left[(u + 2\pi\mu)\tau\right]\ e^{-u^2/2\sigma_{\mathbf{f}}^2}\ du. \tag{4.23}$$

Keeping in mind that the cosine term in *Equation (4.23)* can be expanded such that

$$\cos(A \pm B) = \cos A \cos B \mp \sin A \sin B, \tag{4.24}$$

Equation (4.23) can be expanded to

$$\begin{aligned} E\left[\cos(2\pi \mathbf{f}\tau)\right] \quad = \quad & \frac{1}{\sqrt{2\pi}\sigma_{\mathbf{f}}}\left[\cos(2\pi\mu\tau)\int_{-\infty}^{\infty}\cos(u\tau)\ e^{-u^2/2\sigma_{\mathbf{f}}^2}\ du - \right. \\ & \left. \sin(2\pi\mu\tau)\int_{-\infty}^{\infty}\sin(u\tau)\ e^{-u^2/2\sigma_{\mathbf{f}}^2}\ du\right]. \end{aligned} \tag{4.25}$$

Using the definite integral solutions

$$\int_0^{\infty} \cos bx\ e^{-ax^2}\ dx = \frac{1}{2}\sqrt{\frac{\pi}{a}}\ e^{-b^2/4a} \qquad [a > 0], \tag{4.26}$$

and

$$\int_{-\infty}^{\infty} \sin bx\ e^{-ax^2}\ dx = 0, \tag{4.27}$$

Equation (4.25) now reduces to

$$E\left[\cos(2\pi \mathbf{f}\tau)\right] = \cos(2\pi\mu\tau)\ e^{-(\tau\sigma_{\mathbf{f}})^2/2}. \tag{4.28}$$

The autocorrelation of *Equation (4.28)* is now of a periodic nature rather than just an exponentially decaying function. The autocorrelation function based on *Equation (4.12)* and *Equation (4.28)* can now be written as

$$\psi(\tau) = \frac{1}{2}\sigma_{\mathbf{r}}^2 \cos(2\pi\mu\tau)\ e^{-(\tau\sigma_{\mathbf{f}})^2/2}. \tag{4.29}$$

To see why the derivation of the autocorrelation function is so important it has to be realized that there is one problem with the stochastic model in *Equation (4.2)* concerning this implementation. In order to obtain the autocorrelation function *Equation (4.17)* and *Equation (4.18)* by inserting the appropriate distributions into *Equation (4.2)*, at least one distribution inserted is non-white. In *Subsection 1.4.3*, one of the requirements of the shaping filter is that a white-noise source has to drive the shaping filter. This requires two steps.

First, the white-noise source random variable $\mathbf{x}$ has to be converted to a uniform random variable by

$$\mathbf{u} = F_x(\mathbf{x}), \tag{4.30}$$

where F_x is the cumulative distribution function (CDF) of the random variable $\mathbf{x}$. Then the uniform random variable $\mathbf{u}$ in the interval (0,1) needs to be transformed to the desired random variable $\mathbf{y}$ with known CDF F_y by letting

$$\mathbf{y} = F_y^{-1}(\mathbf{u}), \tag{4.31}$$

Thus, in order to realize the random process given in *Equation (4.2)*, shaping filters transforming all non-white random variables to white ones would be required in the configuration shown in *Figure 4.2*.

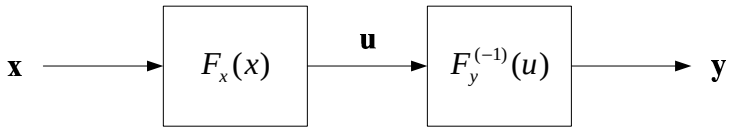

Figure 4.2: Systematic process required to convert non-white-noise source x to white-noise source y.

However, a better alternative exists. *Equation (4.18)* can be represented by a special form of the *Gauss-Markov* process as will be shown in *Chapter 6*. The Gauss-Markov is a n^{th}-order process that produces an autocorrelation function similar to *Equation (4.18)* when driven by a white-noise process.

The speech signal has been defined (*Subsection 4.3.1*) and a stochastic model derived (*Subsection 4.3.2*). The next logical functional unit that needs to be designed in the model of the auditory system in *Figure 4.1* is the signal decomposition functional unit.

4.4 SIGNAL DECOMPOSITION MODEL

The signal decomposition model in *Figure 4.1* has the task to perform a similar function to that of the basilar membrane and the nerve fibers innervating the inner haircells. From *Chapter 2* it is evident that the travelling wave of the basilar membrane vibrates with a maximum amplitude at a location along the cochlea that is dependent on the frequency of stimulation. Each nerve fiber is most sensitive to a particular frequency and the individual fibers within the auditory nerve are also organised according to the particular frequency at which they are most sensitive and the location along the cochlea at which they innervate haircells. The auditory nerves are thus organised topographically, which is known as *tonotopic organisation* or *place theory*.

The basilar membrane as such can thus be seen as a spectrum analyser spreading the information of the spectrum of the signal across a large number of neural channels. This is the primary function of the signal decomposition model. But before such a design can be made,

an in-depth understanding of how a real signal with amplitude, frequency and phase can be spilt in the frequency domain, needs to be gained.

4.4.1 Signal Decomposition Fundamentals

In order to be able to successfully break a signal into its most elementary components, knowledge of the fundamental signal characteristics is of importance. Any sinusoid can be specified completely by its amplitude, frequency and phase:

$$x(t) = A\sin(2\pi ft + \phi). \tag{4.32}$$

For any periodic signal there exists a unique *Fourier series* (Ziemer, Tranter and Fannin, 1998), which for the trigonometric case is given by

$$x(t) = a_0 + \sum_{n=1}^{\infty} a_n \cos(2\pi n f_0 t) + \sum_{n=1}^{\infty} b_n \sin(2\pi n f_0 t), \tag{4.33}$$

where the cosine and sine terms are called the *basis function*, a_0 the *time average* and f_0 the fundamental frequency. *Equation (4.33)* implies that every periodic signal can be represented by a sum of harmonically related sinusoidal terms. In the frequency domain the periodic signal can be represented as frequency components at $0, \pm f_0, \pm 2f_0 \ldots$, which make up the spectrum.

At this point it should be obvious to the reader that one method of extracting all information required to successfully reconstruct a periodic signal with an amplitude, frequency and phase is to extract the harmonically related sinusoidal terms. One option that follows logically from a Fourier series is to divide the spectrum in frequency bands centred around the frequency components $0, \pm f_0, \pm 2f_0 \ldots$. The frequency resolution is dependent on the observation window in the time domain. By increasing the observation window the frequency resolution increases but the temporal resolution decreases. For example, knowing that formant frequencies are only static for a period of approximately 20 ms and assuming the input signal to be a pure tone situated at 300 Hz, the main-lobe width of the signal in the frequency domain would be 100 Hz allowing for no greater frequency resolution than 100 Hz. By increasing the observation window the frequency resolution would increase but spectral "smearing" would occur, a term given to changes in amplitude, frequency or phase taking

place during the observation window period.

Nevertheless, dividing the spectrum in frequency bands can be done by applying a filter bank to the incoming signal, where each filter occupies a specific frequency band. The result is a group of frequency components that can now be transmitted separately and combined at any stage in terms of a Fourier series to represent the original signal.

4.4.2 Filter Banks

Most common spectral analysis/synthesis models in auditory signal processing applications utilise a bank of filters closely related to the filtering properties of the basilar membrane such as the *gammatone* and *gammachirp* filters (Irino and Unoki, 1999). The gammachirp filter is an extension of the popular gammatone filter, where the gammachirp filter extends the symmetric amplitude response of the gammatone filter to an asymmetric one. *Figure 4.3* shows the frequency response of the basilar membrane closely modelled by a gammachirp filter.

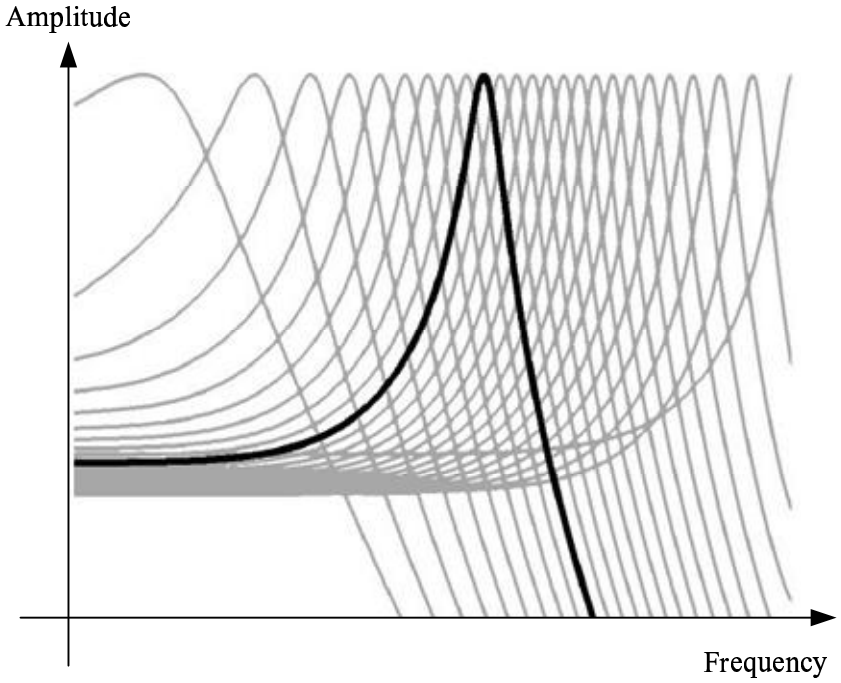

Figure 4.3: Frequency response of points along basilar membrane. Reproduced from Irino & Unoki (1999).

The gammachirp filter bank extends logarithmically in frequency, which can be closely linked

to logarithmic tonotopic organisation in the cochlear (Yost, 2000). However, in order to keep the signal decomposition functional unit as generic as possible the spectral bands of the filter banks in this dissertation are assumed to be linearly spaced in frequency and symmetrical around the centre frequencies. The individual pass-bands are identical in shape and can be thought of as a base-band filter being translated up in the frequency domain to the required centre frequencies.

Another study that supports a filter bank theory as far as signal decomposition is concerned is the findings by Rieke et al. (1997) that auditory neurons tuned to high frequencies are sensitive to fluctuations in the envelope of the waveform as seen through a bandpass filter. There is thus enough evidence supporting the theory that the basilar membrane response combined with particularly tuned neurons, performs the function of an analysis filter bank.

4.4.3 Signal Decomposition Tools

In order to model the spectral analysis function of a bandpass filter bank different techniques can be used. Given below are some of the tools available. Popular transforms are the fast Fourier transform (FFT) (Haykin, 1994; Ziemer et al., 1998; Oppenheim and Schafer, 1989), fast Hartley transform (Le-Mgoc and Vo, 1989), direct cosine transform (DCT) (Rao and Yin, 1990; Watkinson, 1995), Chirp-z transform (Ziemer et al., 1998) and wavelet decomposition (Vetterli, 1992; Daubechies and Maes, 1996; Walker, 1999). The input signal can also be directly filtered by an infinite impulse response filter bank or finite impulse response filter bank. The FFT is the most popular method for spectral analysis due to its fast and efficient algorithm. DCTs are popular in video image encoding seeing that returns double the frequency resolution than that of a discrete Fourier transform (DFT). In this case the infinite impulse response (IIR) filter bank approach was chosen because of its infinite nature that can be easily implemented by a shaping filter as will become evident in *Section 4.6* and *Chapter 6*.

4.4.4 System Decomposition Model

The auditory system model including the source and analysis filter bank is depicted in *Figure 4.4*.

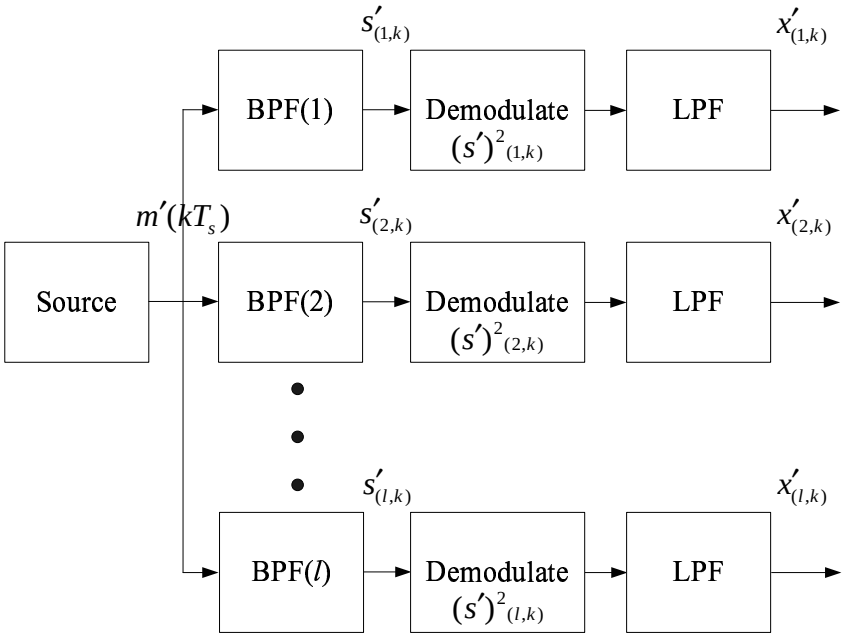

Figure 4.4: Signal decomposition diagram.

The signal generated by the source, defined in *Subsection 4.3.1*, is passed through a bank of l frequency selective filters that subdivides the spectrum into equal but overlapping frequency bands, which are assigned to specific neural channels. The frequency selective information is then demodulated and low-pass filtered as motivated in *Section 4.5*, to extract the power contents of the signal. The bandpass filter bank is implemented by a bank of IIR filters as is discussed below.

4.4.5 Infinite Impulse Response Filter

The infinite impulse response (IIR) filter can be written in the z-domain as

$$H(z) = \frac{\sum_{n=0}^{M} b_n z^{-n}}{1 - \sum_{n=1}^{N} a_n z^{-n}}, \tag{4.34}$$

and as the difference equation

$$y[k] - \sum_{n=1}^{N} a_n y[k-n] = \sum_{n=0}^{M} b_n x[k-n]. \tag{4.35}$$

The coefficients a_n represent the delay terms in *Equation (4.35)* and determine the poles of a linear time-invariant system. The coefficients b_n on the other hand determine the zeros of the linear time-invariant system. It is evident from *Equation (4.35)* that $H(z)$ is not of finite length in the sampled time domain. In this class of filters at least one non-zero pole is not cancelled by a zero. The primary advantage of IIR filters over finite impulse response (FIR) filters is that they typically meet a given set of specifications with a much lower filter order than a corresponding FIR filter. IIR filters unlike their higher-order counter parts are non-linear phase filters, which make them unsuitable for any linear-phase applications. Classical IIR filter types are Butterworth, Chebyshev Types I and II, elliptic and Bessel filters.

One of the greatest distinguishing factors between the aforementioned IIR filters and the elliptic filter, which implements the filter bank in this dissertation, is the filter order required to obtain a specified response. It also offers a steeper roll-off than the Butterworth or Chebychev I, II filter. Its drawback is that it is equiripple in both the pass and stopband.

The IIR filter due to its infinite transfer function length can be thought of as a p^{th} order autoregressive (AR), moving average (MA) process, or ARMA process. The ARMA process is discussed in *Chapter 6* and its relevance to this IIR filter bank will become clear.

In this dissertation the bandpass filter bank was implemented by means of IIR filters. The reason for choosing IIR filtering above any other signal decomposition tool mentioned in

Subsection 4.4.3 is due to its two main advantages:

1. The IIR filter typically meets a given set of specifications with a lower filter order than FIR filters and the *elliptic* filter produces the lowest order implementation of any type of filter and

2. the IIR implementation is closely related to that of a shaping filter.

The first advantage stated is quite straightforward and the reader is referred to Oppenheim and Schafer (1989) for more detail on the filter design. The last advantage stated is however less clear and will be discussed in more detail below.

In any Kalman filtering problem the internal model of the Kalman filter needs to statistically mimic the behaviour of the underlying process in order for the Kalman filter to be the best optimal estimator for the problem. Usually the underlying process is subjected to a system identification (SID) procedure (Gelb, 1984) from which a process can be derived that mimics the behaviour of the process and fulfills the requirements of an internal model for the Kalman filter. As was mentioned in *Section 1.4*, the real process needs to be modelled by a shaping filter. In this dissertation, an alternative approach to SID or power spectrum estimation is taken. The IIR filter bank is directly converted to a shaping filter. Not all filters however can be directly translated into a shaping filter. Below, it will be shown why the FIR filter is not a candidate for direct translation but the IIR is.

The discrete-time shaping filter can be derived from the continuous shaping filter given in *Chapter 1*, *Figure 1.7* and is shown in *Figure 4.5*.

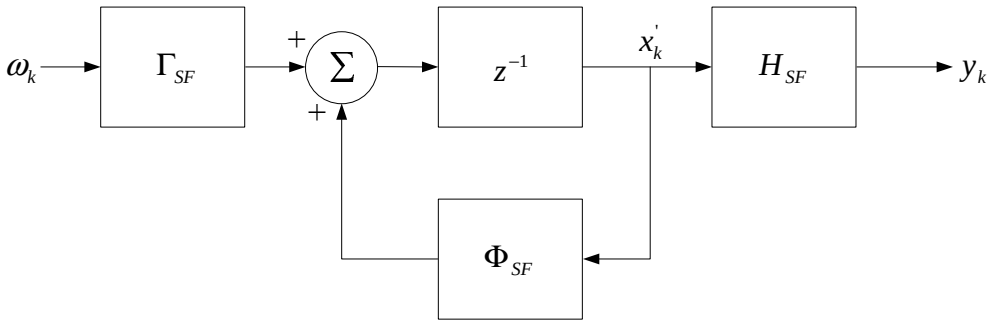

Figure 4.5: Special form of discrete-time shaping filter.

In the sampled time domain the recursive shaping filter equations are given by:

$$x_{k+1} = \Phi_{SF} x_k + \Gamma_{SF} \omega_k, \tag{4.36}$$

and

$$y_k = H_{SF} x_k. \tag{4.37}$$

To simplify the problem it will be assumed for the moment that all matrices in *Figure 4.5* are scalars. It is evident from *Equation (4.36)* that the input ω_k is coupled by a variable Γ_{SF} to the recursive loop. Φ_{SF} modifies the previous state estimate x_k and feeds it back to the coupled input. Keeping the difference equation *Equation (4.36)* in mind a look is now taken at the FIR and IIR filters.

The FIR filter has, as its name suggests, an impulse response of finite length and can be written in the z-domain as

$$H(z) = \sum_{n=0}^{M} b_n z^{-n}, \tag{4.38}$$

or as a convolutional sum in the sampled time domain as

$$y\left[k\right] = \sum_{n=0}^{M} b_n x\left[k-n\right], \tag{4.39}$$

where $x[k]$ is an input. The IIR on the other hand as discussed in *Subsection 4.4.5* has an infinite impulse response in the z-domain given by

$$H(z) = \frac{\sum_{n=0}^{M} b_n z^{-n}}{1 - \sum_{n=1}^{N} a_n z^{-n}}, \tag{4.40}$$

and a difference equation given by

$$y[k] - \sum_{n=1}^{N} a_n y[k-n] = \sum_{n=0}^{M} b_n x[k-n]. \tag{4.41}$$

If it is assumed for a moment that H_{SF} is simply one and the output is equal to the state variable x_n then it is apparent from *Equation (4.39)* that no feedback mechanism is present in the case of the FIR filter. The IIR filter on the other hand possesses a feedback mechanism through its numerator coefficients a_n (see *Equation (4.41)*). Thus, the first requirement of transforming any filter to a shaping filter is that poles are present. In fact as is discussed in *Chapter 6* the order of the polynomial in the denominator has to be at least one more than the numerator polynomial. In *Chapter 6* the expansion from a scalar to matrix format is discussed and how both numerator and denominator coefficients are accommodated in the shaping filter by means of the *observable canonical form*, the *controllable canonical form* and others.

In the next section it is illustrated how an IIR filter can be implemented by an ARMA model.

4.4.5.1 ARMA model implementing a filter

An ARMA model can implement a transfer function of the form

$$H(z) = \frac{D(z)}{A(z)} = \frac{\sum_{n=0}^{q} d(n) z^{-n}}{1 + \sum_{n=1}^{p} a(n) z^{-n}}. \tag{4.42}$$

ARMA models are ideal for implementing rational and realisable transfer functions in state-space form, something that is desired when designing the Kalman filter. They can also be used in system identification problems, where an unknown process with an experimentally determined autocorrelation function needs to be represented in state-space form. ARMA models can however represent only stationary time series, something that is avoided by using an autoregressive integrated moving average (ARIMA) process, of which the random-walk process is a simple example.

Given the transfer function *Equation (4.42)*, if the number of coefficients in the numerator (q) and denominator (p) are equal ($p = q$), where the denominator is normalised to one, the transfer function *Equation (4.42)* is said to be *proper* and can be written in state-space form as (Furuta, Sano and Atherton, 1988; Burghes and Graham, 1980; DeRusso, Roy and Close, 1965)

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \\ x_3(k+1) \\ \vdots \\ x_p(k+1) \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & \dots & -a_p \\ 1 & 0 & 0 & \dots & -a_{p-1} \\ 0 & \ddots & \vdots & \vdots & \vdots \\ 0 & \dots & 1 & 0 & -a_2 \\ 0 & 0 & \dots & 1 & -a_1 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \\ x_3(k) \\ \vdots \\ x_p(k) \end{bmatrix} + \begin{bmatrix} d_1 - a_1 d_0 \\ d_2 - a_2 d_0 \\ d_3 - a_3 d_0 \\ \vdots \\ d_p - a_p d_0 \end{bmatrix} u(k), \tag{4.43}$$

and

$$y(k) = \begin{bmatrix} 0 & \dots & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \\ \vdots \\ x_p(k) \end{bmatrix} + d_0 u(k). \tag{4.44}$$

However, if the number of denominator coefficients (p) is one more than the number of numerator coefficients (q) so that $q = p - 1$ then the direct feed-through that exists from the input to the output via the coefficient d_0 in (4.44) does no longer exist and the transfer function driven by the source $u[k]$ is said to be *strictly proper* and can be written in the discrete-time state-space form

$$\begin{aligned} \mathbf{x}[k+1] &= \mathbf{A}\mathbf{x}[k] + \mathbf{B}u[k], \\ y[k] &= \mathbf{C}\mathbf{x}[k], \end{aligned} \tag{4.45}$$

where $b_i = d_i$. The feed-forward matrix $\mathbf{D}$ in this case is zero, which is significant since the state-space form of a shaping requires that the feed-forward matrix be either zero or be reserved for the measurement noise process (see *Figure 1.7* and *Equation (1.1)*), at least in its

most fundamental form.

If a filter has the required number of numerator and denominator coefficients then the filter coefficients b and a can be implemented in the *observable canonical form* where

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 0 & \dots & -a_{p-1} \\ 1 & 0 & 0 & \dots & -a_{p-2} \\ 0 & \ddots & \vdots & \vdots & \vdots \\ 0 & \dots & 1 & 0 & -a_2 \\ 0 & 0 & \dots & 1 & -a_1 \end{bmatrix}, \tag{4.46}$$

$$\mathbf{B} = \begin{bmatrix} b_1 & b_2 & b_3 & \dots & b_{p-1} \end{bmatrix}^T, \tag{4.47}$$

and

$$\mathbf{C} = \begin{bmatrix} 0 & 0 & 0 & \dots & 1 \end{bmatrix}. \tag{4.48}$$

By inserting *Equation (4.46)* to *Equation (4.48)* in the state-space equation, *Equation (4.45)*, and driving it with a source $u[k]$ the filter is applied by means of an ARMA process, which acts like a shaping filter.

Most systems are however not found in a *strictly proper* form since the number of poles and zeros, which are part of the filter design process do not always result in the desired form. Designing, for example, a digital IIR filter in MATLAB, results in a proper transfer function rather than the desired strictly proper transfer function with no direct feed-through. In order to convert the transfer function, the number of coefficients in the denominator has to be one more than the numerator. Given a denominator polynomial of the form

$$z^0 + a_1 z^{-1} + a_2 z^{-2} + \ldots + a_{n-1} z^{1-n} + a_n z^{-n}, \tag{4.49}$$

then by shifting the coefficients to the left by one and inserting a zero at z^{-n} the polynomial becomes

$$z^1 + a_1 z^0 + a_2 z^{-1} + \ldots + a_n z^{1-n} + 0z^{-n}. \tag{4.50}$$

This mathematical manipulation is nothing more than adding an extra pole to the system, increasing the order of the system by one.

The transition from proper to strictly proper is illustrated with the following example. Given a fifth-order lowpass elliptic IIR filter with a cutoff frequency of 220 Hz and a sampling frequency of 600 Hz, the transfer function is given by

$$\frac{0.2745z^5 + 1.198z^4 + 2.244z^3 + 2.244z^2 + 1.198z + 0.2745}{z^5 + 2.035z^4 + 2.401z^3 + 1.422z^2 + 0.5243z + 0.0505}. \tag{4.51}$$

The magnitude and phase responses are shown in *Figure 4.6*.

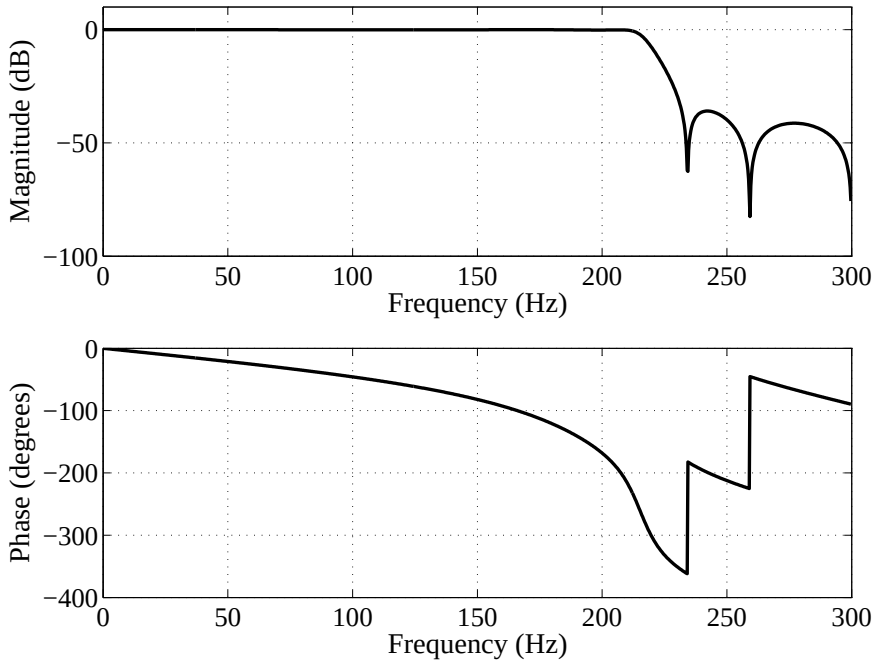

Figure 4.6: Magnitude and phase response plot.

The pole/zero locations are depicted by *Figure 4.7*.

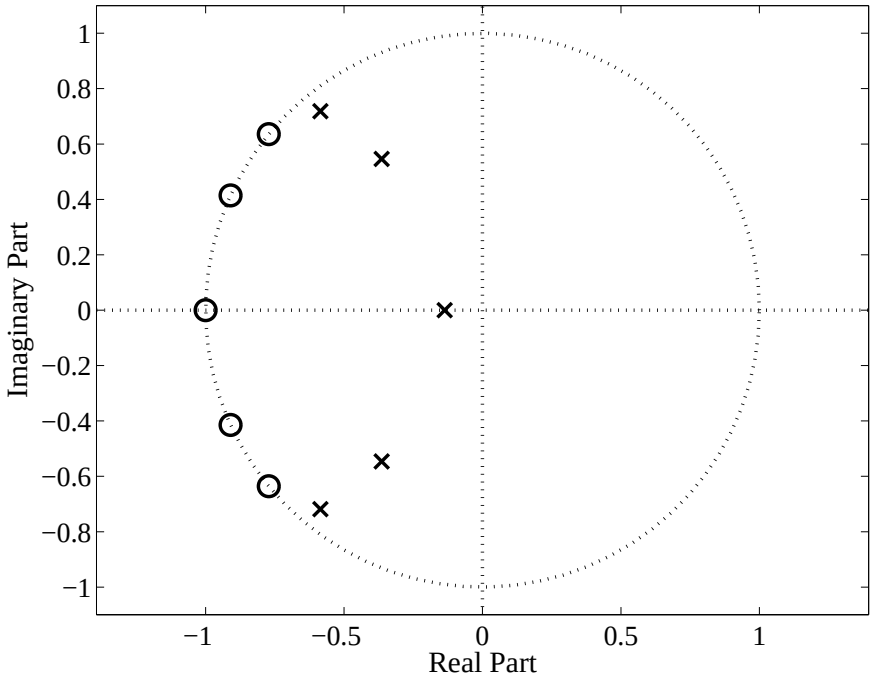

Figure 4.7: Pole/zero plot of fifth-order elliptic LPF in proper form.

The filter can also be represented in state-space format as:

$$\begin{aligned} \mathbf{x}[k+1] &= \mathbf{A}\mathbf{x}[k] + \mathbf{B}u[k], \\ y[k] &= \mathbf{C}\mathbf{x}[k] + \mathbf{D}u[k], \end{aligned} \tag{4.52}$$

where

$$\mathbf{A} = \begin{bmatrix} -2.035 & -0.600 & -0.178 & -0.066 & -0.013 \\ 4.000 & 0 & 0 & 0 & 0 \\ 0 & 2.000 & 0 & 0 & 0 \\ 0 & 0 & 1.000 & 0 & 0 \\ 0 & 0 & 0 & 0.500 & 0 \end{bmatrix}, \tag{4.53}$$

$$\mathbf{B} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \end{bmatrix}^T, \tag{4.54}$$

$$\mathbf{C} = \begin{bmatrix} 0.639 & 0.397 & 0.232 & 0.132 & 0.065 \end{bmatrix} \tag{4.55}$$

and

$$\mathbf{D} = 0.065. \tag{4.56}$$

By adding an extra pole to the system the transfer function of *Equation (4.51)* becomes

$$\frac{0.2745z^5 + 1.198z^4 + 2.244z^3 + 2.244z^2 + 1.198z + 0.2745}{z^6 + 2.035z^5 + 2.401z^4 + 1.422z^3 + 0.5243z^2 + 0.0505z}. \tag{4.57}$$

The magnitude and phase plot of *Figure 4.6* stay the same but as the pole/zero plot of *Figure 4.8* illustrates an additional pole is found at zero.

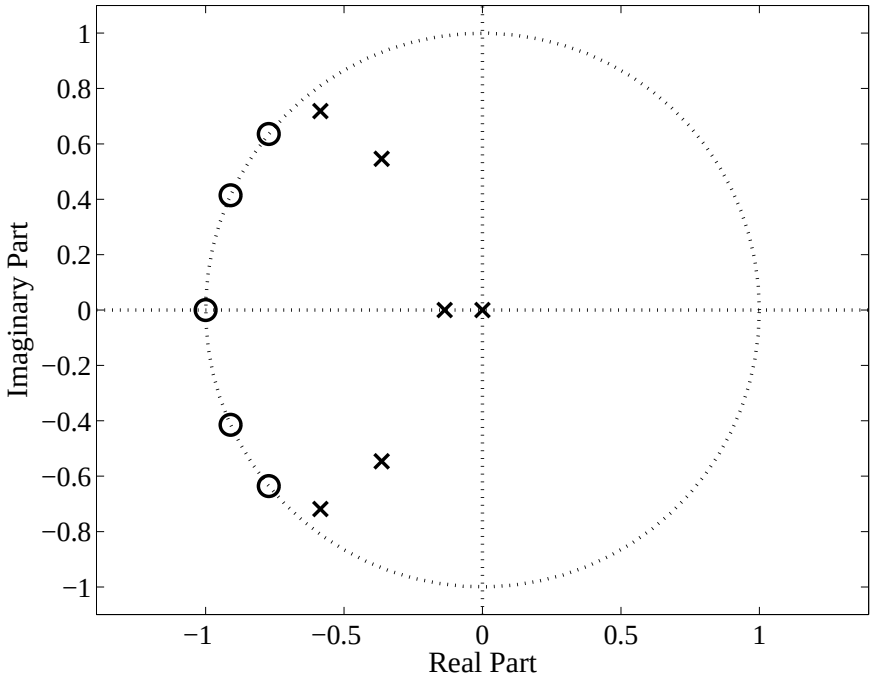

Figure 4.8: Pole/zero plot with additional pole at zero.

Rewriting the transfer function *Equation (4.51)* into a state-space format results in

$$\mathbf{A} = \begin{bmatrix} -2.035 & -0.600 & -0.178 & -0.033 & -0.003 & 0 \\ 4.000 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2.000 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2.000 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1.000 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.250 & 0 \end{bmatrix}, \tag{4.58}$$

$$\mathbf{B} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}^T, \tag{4.59}$$

$$\mathbf{C} = \begin{bmatrix} 0.275 & 0.299 & 0.281 & 0.140 & 0.075 & 0.069 \end{bmatrix} \tag{4.60}$$

and

$$\mathbf{D} = 0. \tag{4.61}$$

The matrix $\mathbf{D}$ is as expected now zero and does not feed the input directly to the output as required by a strictly proper system of which the shaping filter is one.

An alternative method of finding a suitable transfer function describing the underlying process is spectrum estimation discussed below.

4.4.5.2 Spectrum estimation as an alternative method

Usually the problem of mapping any filter to a shaping filter is not done by direct mapping as was described in *Chapter 4* due to the problems that were encountered, such as filter coefficients not directly fitting into an ARMA process. Many Kalman filter designers do not have the advantage of knowing the exact transfer function of the system they are trying to track, since it may be unknown to them, for one or other reason. They then often revert to techniques such as system identification by means of spectral estimation to find an equivalent transfer function. Most of the shaping filter designs are done by using one or other power

spectrum estimation technique as discussed in *Addendum A.1*.

In *Figure 4.4* the signal after being passed through a bank of IIR filters is demodulated and then low-pass filtered to find the power in each band. The power calculation principle is described next.

4.5 POWER CALCULATION PRINCIPLE

It could be argued that in the auditory system, once the information has been split into spectral bands, the amount of power or energy present within each band specifies the *intensity* at a specific position within the spectral band. Seeing that the position is known due to a tonotopic organisation of the neural channel, a *power spectral map* discussed briefly in *Section 1.4* can be found. The resolution of the power spectral map is determined by the width of the bandpass filters and forms part of the fundamental low-level information that is available to the CANS. In this case no interest lies with how the information is used by the CANS or even the CNS but more in how information can be coded *reliably* in such a configuration.

In order to obtain a scaled version of the power or energy, the signal coming from each bandpass filter can be squared. This process can be seen as *demodulating* the band-passed signal as will be shown next.

4.5.1 Demodulation

4.5.1.1 Perfect demodulation of all frequencies

In AM demodulation it is generally desired to demodulate a message signal $m(t)$ with a spectrum centred around the carrier frequency f_c, to baseband. In this case, the power in a certain part of the spectrum as specified by the width of the BPF has to be estimated. A message signal $m(t)$ modulated by a carrier frequency f_c can be given by

$$s(t) = A_c m(t) \cos(2\pi f_c t), \tag{4.62}$$

where the spectrum is given by

$$S(f) = \frac{A_c}{2}\left[M(f - f_c) + M(f + f_c)\right]. \tag{4.63}$$

The demodulated signal is of the form (see *Addendum A.2* for derivation).

$$d(t)_{AM} = A_c A_c' \frac{m(t)}{2}\left[\cos\phi + \cos(4\pi f_c t + \phi)\right]. \tag{4.64}$$

This process, called AM demodulation, demodulates a band centred around f_c, to baseband. The demodulated signal is then low-pass filtered to remove the high frequency component found in *Equation (4.64)*.

If a message signal $m(t)$ is however demodulated by squaring the signal a slight alteration of *Equation (4.64)* is required, resulting in

$$d(t)_{SQR} = A_c A_c' \frac{m(t)^2}{2}\left[\cos\phi + \cos(4\pi f_c t + \phi)\right]. \tag{4.65}$$

When a sampled modulated signal is squared the demodulation process can be thought of as the demodulation of each frequency component with itself and the other frequency components in the band, which has the effect of demodulating every component of the spectrum to the DC component with some additional unwanted components at other non-zero frequencies. Thus all power that was found in the message signal is now found as a cumulative sum at DC. However, for each frequency component a scaled version is found at twice the carrier frequency as *Equation (4.65)* indicates. Cross-modulated frequency terms are also generated by squaring the message signal. In order to obtain only the DC component, the demodulated signal has to be low-pass filtered with a very narrow low-pass bandwidth. Some residue of the unwanted frequency components is to be expected in the low-passed signal, something that is investigated in more detail in *Section 4.6*.

4.5.1.2 Remarks on conversion from power units to Volt

In *Section 4.3* the amplitude of the input signal was defined as being measured in volts. After demodulation takes place the units of the demodulated signal becomes volts squared (i.e.

power if measured over a 1 Ω resistor). However, in *Chapter 5* it is required that the leaky integrate-and-fire neuron be driven by a signal with units in volts. Since it was desired to drive the LIF neuron directly by the demodulated signal, the magnitude of the demodulated signal had to be converted to a value in volts. Thus, the relationship between the demodulated signal and the LIF input is linear with a gain of one.

4.6 SIMULATION

The aim of this section is to illustrate exactly how a speech signal defined in *Section 4.3* can be generated and then decomposed as shown in *Figure 4.4*. The speech model is evaluated and the power spectral density (PSD) found, which will be compared to a stochastic model found in *Chapter 6*.

4.6.1 Speech Signal Implementation

The quasi-periodic signal with random amplitude, frequency and phase was implemented with the following specifications:

1. The random amplitude is a Gaussian RV with $N(0\text{ V}, 1/6\text{ V}^2)$,
2. where the random frequency is a Gaussian RV with $N(307\text{ Hz}, 25\text{ Hz}^2)$ and
3. the random phase is a uniform RV with zero mean on the interval $(-\pi, \pi)$.

The signal jumps to the next random frequency every $t_{skip} = 20$ ms. The amplitude was limited to have a maximum value of one by clipping all amplitudes above one and the frequency variation was kept small in order to avoid too much clipping taking place, which would change the statistics of the simulated signal considerably. The particular choice for a mean of 307 Hz, for the source signal will become apparent in the next few sections. The signal was generated by amplitude modulating the random amplitude with a random frequency and phase as specified above. Every 20 ms a new amplitude, frequency and phase was chosen from the specified distributions.

To generate the speech-like signal of a quasi-periodic nature, principles from AM were used. AM can be thought of as modulating a signal $m(t)$ with a carrier f_c

$$c(t) = A_c \cos(2\pi f_c t). \tag{4.66}$$

The amplitude modulated wave can be described in the most general form by

$$s(t) = A_c \left[1 + k_a m(t)\right] \cos(2\pi f_c t), \tag{4.67}$$

where f_c is the carrier frequency, A_c the carrier amplitude, k_a the amplitude sensitivity and $m(t)$ message signal. By choosing

$$|k_a m(t)| < 1, \tag{4.68}$$

the function

$$1 + k_a m(t) > 0, \tag{4.69}$$

avoiding overmodulation and phase reversal.

Once the message signal has been amplitude modulated it is passed through the IIR filter bank. By implementing the speech signal the following results can be obtained.

4.6.1.1 Speech signal implementation results

The source signal having the characteristics of a speech signal is simulated and plotted in the time and frequency domain. The autocorrelation is also determined and plotted in terms of time differences between samples (τ).

Figure 4.9 depicts the generated random source signal as a function of time, taken over an arbitrary 1 s interval.

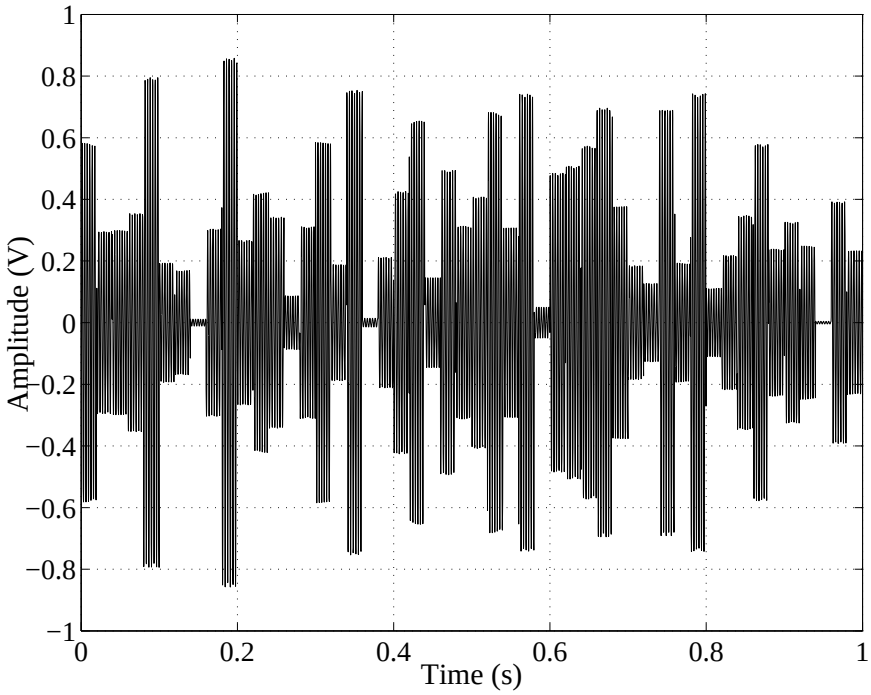

Figure 4.9: Random speech-like source signal.

The quasi-periodic signal has a constant amplitude and frequency over a period of 20 ms at a time, where the magnitude of the signal is never larger than 1. The signal has the appearance of that of a speech signal because of its quasi-periodic nature and its clearly distinct fixed-interval formant frequencies. Unlike a true speech signal (Cole, 1980) no drifting of formant frequencies occurs and only one formant is present at a time.

In *Figure 4.10* the PSD is shown as a function of frequency from 0-700 Hz.

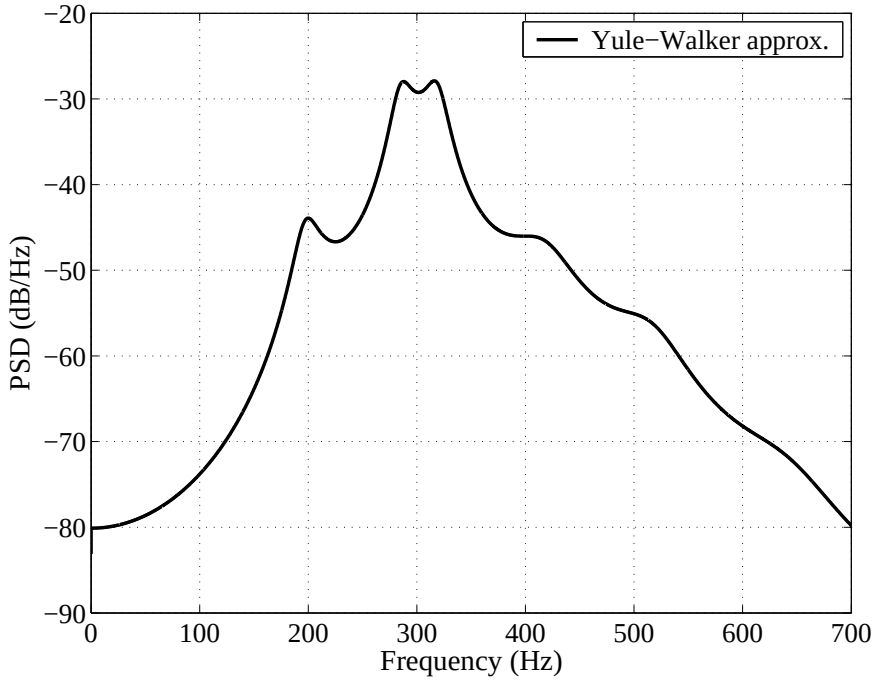

Figure 4.10: PSD of source signal.

The Yule-Walker (Marple, 1987; Kay, 1988) algorithm was used to estimate the PSD of the random process. The power has the approximate form of a Gaussian distribution, centred roughly around 300-310 Hz with a peak average power of roughly -28.0 dB/Hz over an interval of 50 s.

The PSD is a clear indication of the average formant frequency behaviour of the generated speech signal over an arbitrary 50 s interval. Over a 50 s interval the formant had changed 2500 times and was primarily located around the centre frequency of 307 Hz as defined in *Subsection 4.6.1*

The normalised autocorrelation function of the random input signal is shown in *Figure 4.11*.

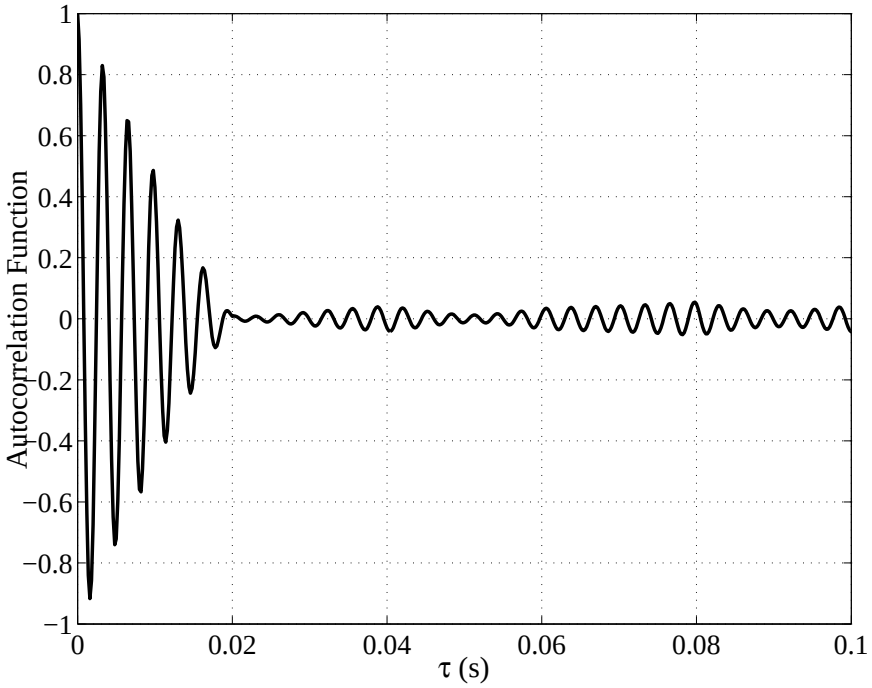

Figure 4.11: Normalised autocorrelation function of input signal.

The autocorrelation function of the simulated signal exhibits a periodic behaviour, where the peaks are situated at ± 3.25 ms, ± 6.5 ms and ± 9.75 ms from $\tau = 0$. The autocorrelation function plot also indicates the quasi-periodic nature of the random input signal over a 20 ms interval. The particulars of the periodic behaviour of the autocorrelation function are required in *Chapter 6* for the design of the Gauss-Markov model.

In the next section the signal extraction functional block is simulated by implementing a bank of elliptic filters.

4.6.2 IIR Filter Implementation

The IIR filter bank was chosen with the following specifications:

1. The number of denominator coefficients (n) is one more ($m = n - 1$) than the number of numerator coefficients (m) and
2. the centre frequencies of filters start at 300 Hz and increase in 15 Hz steps.
3. The filter bandwidths are increased according to a desired overlap rather than a change of centre frequencies,
4. where 0 % overlap corresponds to the -6 dB bandwidth touching the - 6 dB points of the adjacent filters and
5. where 50 % overlap corresponds to a 50 % overlap in the -6 dB bandwidth points of adjacent filters.

4.6.2.1 Filter bank results

In this section the third-order elliptic IIR filter bank for the first three adjacent filters starting at 300 Hz centre frequency representing three neural channels are shown. The filter characteristics are plotted as magnitude and phase plots. Only the first three filters were plotted seeing that the pattern repeats itself for filters lying higher up in the frequency band.

In *Figure 4.12* the frequency response (magnitude plot) is depicted for the case where the filters, as defined, overlap by 0 %.

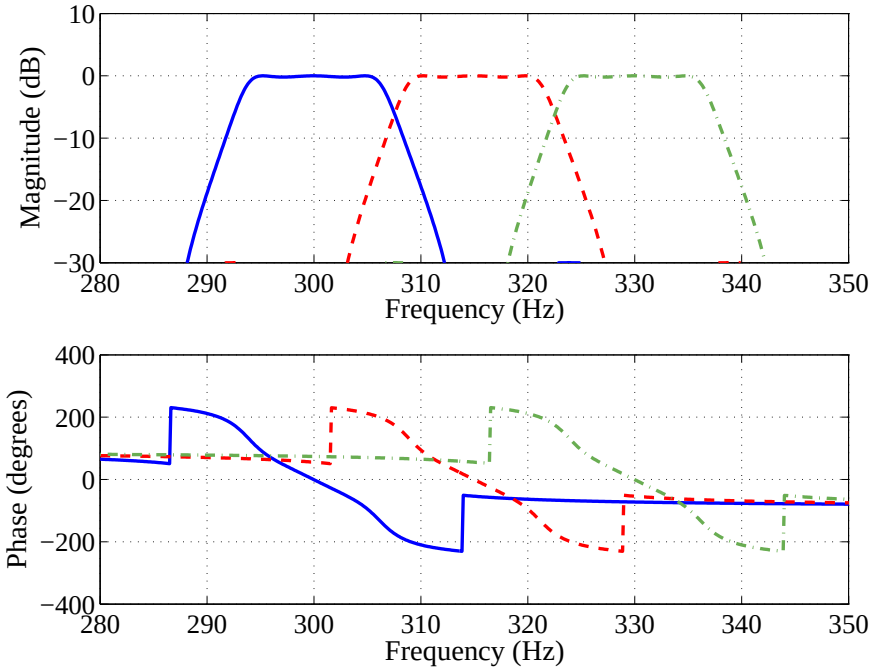

Figure 4.12: Frequency response of third-order elliptic bandpass filters overlapping by 0%.

The three filters are centred at 300, 315 and 330 Hz. Their -6 dB bandwidths are 14.9 Hz and they overlap at 307.5 and 322.5 Hz. The passband ripple is measured to be 0.2 dB and the stopband -30 dB. The phase plot indicates a non-linear phase response.

In *Figure 4.13* the frequency response is depicted for the case where the filters overlap by 50 %.

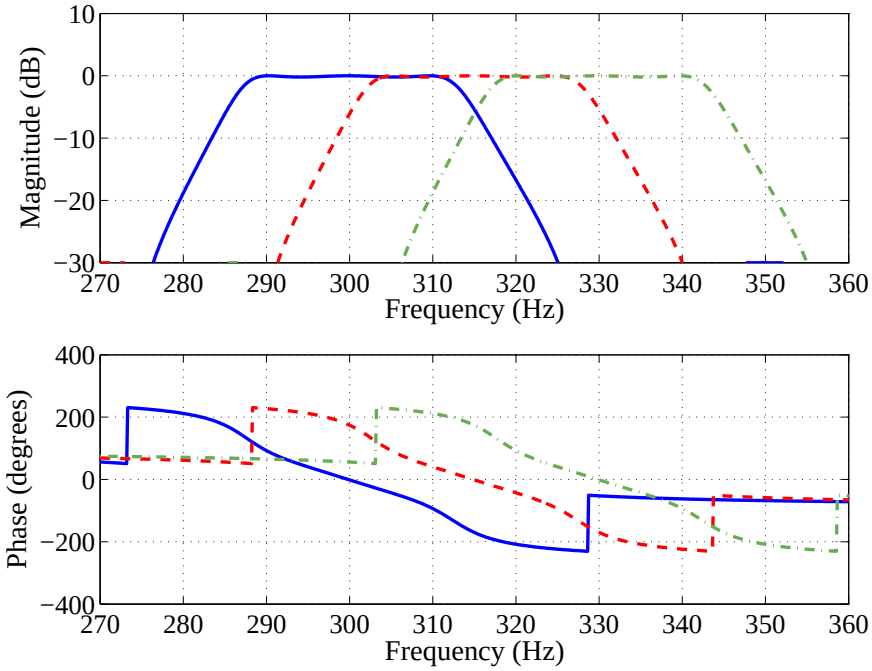

Figure 4.13: Frequency response of third-order elliptic bandpass filters overlapping by 50%.

The three filters are centred as in the case of 0 % overlap at 300, 315 and 330 Hz. The -6 dB bandwidths are however now approximately 30.23 Hz. The -6 dB points of filter 1 (left) are 285.05 and 315.28 Hz, for filter 2 (middle) they are 300.04 and 330.27, and for filter 3 (right) 315.03 and 345.26 Hz. By the definition given in *Subsection 4.6.2* the percentage overlap is calculated to be 50 %. The phase response indicates a non-linear phase. In this dissertation only a single channel implementation will be investigated (50 % overlap).

The filter situated around 300 Hz is used as the default filter with corresponding state-spaces matrices

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & -0.946 \\ 0 & 1 & 0 & 0 & 0 & 0 & 5.105 \\ 0 & 0 & 1 & 0 & 0 & 0 & -12.073 \\ 0 & 0 & 0 & 1 & 0 & 0 & 15.904 \\ 0 & 0 & 0 & 0 & 1 & 0 & -12.297 \\ 0 & 0 & 0 & 0 & 0 & 1 & 5.2968 \end{bmatrix}, \tag{4.70}$$

$$\mathbf{B} = \left(\begin{bmatrix} -3.343 & 11.895 & -13.918 & 0.000 & 13.918 & -11.895 & 3.343 \end{bmatrix} \cdot 10^{-3} \right)^T, \tag{4.71}$$

$$\mathbf{C} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{4.72}$$

and

$$\mathbf{D} = 0. \tag{4.73}$$

In the next section the LPF used for removing any non-DC components as specified in *Subsection 4.5.1* is shown. The results of a simple demodulation procedure in conjunction with the LPF is shown to illustrate how the filter implemented by the MATLAB *filter* function and an ARMA process compare to a theoretically calculated one.

4.6.3 Power Extraction and Low-pass Filtering Results

The extraction of the power of a particular spectral band is obtained in two steps, squaring and low-pass filtering. To compare how well this two-step process is approximate by the simulated process, power from a simple modulated signal is extracted.

The magnitude and phase response of the lowpass filter are shown in *Figure 4.14*.

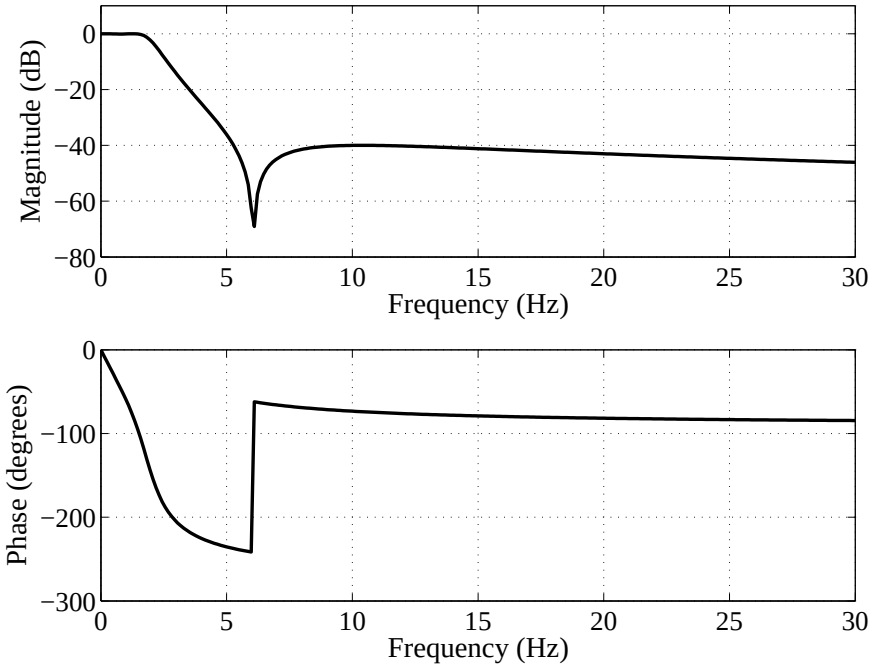

Figure 4.14: Magnitude and phase response of lowpass demodulation filter.

The filter's 3 dB cutoff point is at 2 Hz with a passband ripple of 0.06 dB and a stopband attenuation of 40 dB. The phase response is non-linear.

The state-space equations of the lowpass filter, in *Figure 4.14* are given by

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0.996 \\ 0 & 1 & 0 & -2.991 \\ 0 & 0 & 1 & 2.996 \end{bmatrix}, \tag{4.74}$$

$$\mathbf{B} = \left(\begin{bmatrix} 0.122 & -0.122 & -0.122 & 0.122 \end{bmatrix} \cdot 10^{-3} \right)^T, \tag{4.75}$$

$$\mathbf{C} = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix} \tag{4.76}$$

and

$$\mathbf{D} = 0. \tag{4.77}$$

4.6.3.1 Power extraction results

To demonstrate how the demodulation procedure extracting the power as discussed in *Subsection 4.5* performs in the case of a pure-tone modulated to a carrier frequency, three demodulation *implementations* were compared. The first is a theoretical solution provided by ideal demodulation by means of squaring and then low-pass filtering the signal. The second and third method implement the lowpass filter, in this case an elliptic LPF, by means of the *filter* function in MATLAB and an ARMA process.

The following signal

$$\begin{aligned} s(t) &= m(t)\cos(2\pi f_c t), \\ &= \frac{1}{2}\cos(2\pi 300t), \end{aligned} \tag{4.78}$$

was demodulated and low-pass filtered. Theoretically the power (over a 1 Ω resistor) should be

$$P = \left[\frac{m^2(t)}{2}\right]^2 = 15.625 \text{ mW}. \tag{4.79}$$

In *Figure 4.15* the random input signal of *Figure 4.9*, as observed after the demodulation and low-pass filtering stage, is shown.

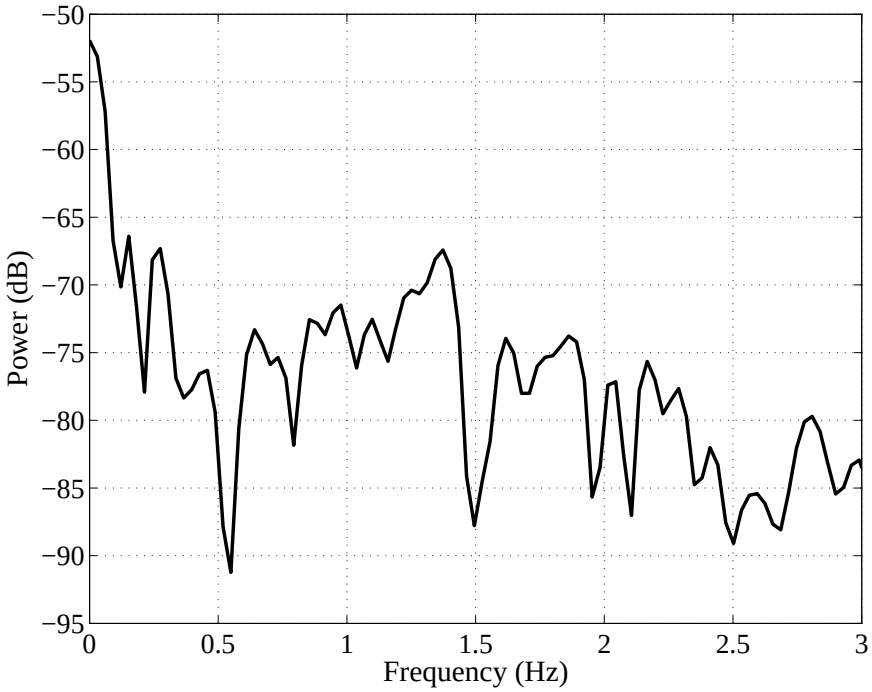

Figure 4.15: Random speech-like signal after demodulation and lowpass filter.

Almost all the power is found within the first 3 Hz of the spectrum, where no significant sidelobes are observed.

The pure sinusoid $0.5\cos(2\pi 300t)$ was demodulated and low-pass filtered. *Table (4.1)* summarises the results:

Table 4.1: Results for pure-tone demodulation.

Model	**Power (mW)**	**Attenuation (dB)**
Theoretically calculated	15.625	-
ARMA	14.132	0.436
MATLAB *filter* function	14.141	0.434

Thus roughly 10 % of the sidelobe power is lost when low-pass filtering the squared signal, where the ARMA model attenuates 0.5 % more than the *filter* function in MATLAB.

4.7 DISCUSSION AND CONCLUSION

In this chapter the source signal and the extraction of bandlimited power by means of a bank of elliptic filters was designed and evaluated.

The random test signal, in this case a speech signal as defined in *Subsection 4.3*, was designed and later implemented in *Chapter 7*. The autocorrelation function given in *Figure 4.11* indicates that a quasi-periodic random process is generated by the source model. The PSD plot shown in *Figure 4.10* shows that the formant frequency indeed changes randomly with time and is centred around 307 Hz with a Gaussian-like distribution. The source signal thus satisfies the design requirements of a formant frequency changing randomly in amplitude, frequency and phase. A matching stochastic signal model for the speech model is found in *Chapter 6*, when the estimator model is designed.

The bandpass and demodulation filter for power extraction purposes meet the specifications given in *Subsection 4.6.2*. The loss of power in the demodulation filter is minimal and can be adjusted by adding an overall gain to the filtered signal. The ARMA process introduces a slightly larger loss than the *filter* function in MATLAB.

Although the bandpass filters were chosen to start at a centre frequency of 300 Hz it is by no means required that they should be positioned at such a frequency. The positions were chosen arbitrarily in order to underline the generic nature of the system. The bandwidths of the bandpass filter banks are also fixed in size suggesting a low frequency resolution across the entire frequency range. Comparing this to the *Weber fraction* (Yost, 2000) where the fraction $\Delta F/F$ is found as a function of frequency, frequency resolutions in the auditory system as small as 1.6 Hz in the range 500 to 2000 Hz have been measured. As was mentioned in *Subsection 4.4.1* allowing for a small bandwidth results in "smearing" of the spectrum being analysed, while a large bandwidth produces only a coarse frequency resolution. This would suggest that the physiology in the auditory system utilises an adaptive filter bank (Irino and Unoki, 1999) changing the spectral resolution as desired. The function of sharpening the frequency selectivity is performed by the outer haircells as was mentioned in *Chapter 2*. However, to keep the system as generic as possible and to illustrate a principle, the fact that a bank of filters is present, rather than to find the exact auditory system parameters for such a filter bank, a small scale model with linearly spaced elliptic filters as simulated in *Subsection 4.6.2* suffices.

Another essential observation made by analysing the power extraction procedure is that the phase information is lost when squaring the bandlimited signal. Since the interest lies only in the magnitude of the power and the frequency information, which is represented in the tonotopic organisation of the filter banks and the neural channels, phase is of no concern here. The information of the speech signal is thus extracted in a generic way, by retaining both power and frequency information.

CHAPTER 5

ENCODER, DECODER AND CHANNEL MODELS

5.1 AIM OF THIS CHAPTER

A model for the source signal and the auditory system, excluding the inner haircells, was designed and simulated in *Chapter 4*. The aim of this chapter as shown in *Figure 5.1* is to design an encoder model representing the mechanical to neural transduction at the inner haircells, a channel and noise model that is linked closely to the biology of the auditory nerve bundle and the noise introduced by unreliable neurons.

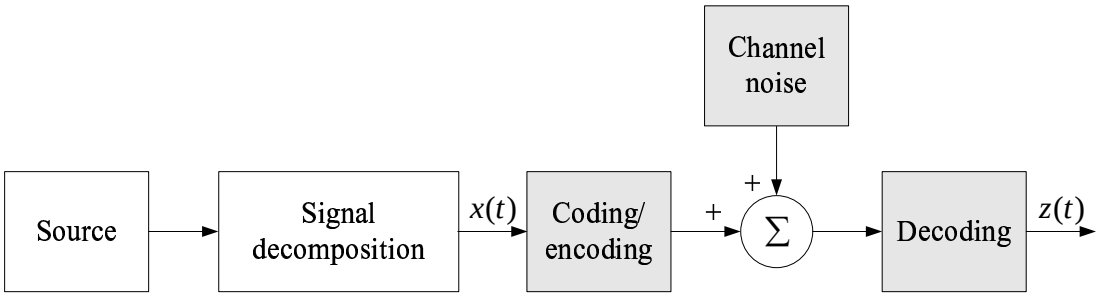

Figure 5.1: The coding/encoding, channel noise and decoding functional units (shaded), are designed and evaluated in this chapter.

A decoder is then designed to estimate the signal at the encoder input in order to evaluate the performance of the encoder/decoder combination in the presence of noise. The encoder, neural channel and decoder are the only parts of the complete system where the physiology is implemented without generalisation. It is the central block, as far as the dissertation is concerned, in that it subjected to a fixed set of constraints commonly found in the biology and physiology.

5.2 INTRODUCTION

The auditory system model found in *Chapter 4* connects to the CANS via the inner haircells. The CANS consists of an intricate network of neurons that transmits neural messages from one location to another as required. As was mentioned in *Chapter 2* an auditory nerve bundle carries the information contained in the spike trains to the CN and from there to the trapezoid body and superior olivary complex. The auditory nerve bundle thus forms a *channel* between the inner haircells and the CANS. The information channel is subjected to noise, to be investigated in this chapter.

In order to model this part of the auditory system, models for the encoder and channel have to be found. Although it is unlikely that signal reconstruction as such takes place in an actual neural system, a similar process is present that in some way or another extracts information from the discrete-time sequence and converts it to a continuous signal. For example the fly (Rieke et al., 1997) can initiate a turn based on a visual motion signal alone encoded as a train of spikes and output an analogue torque signal. The signal reconstruction or decoding in this case aids in evaluating how well the rate code performs if transmitted over a noisy and thus unreliable channel.

Before the encoder and decoder are designed the channel and noise models are defined.

5.3 CHANNEL AND NOISE MODELS

This investigation pivots around the channel model. While the surrounding blocks form an integral part of the system as a whole, the channel and its underlying limitations were closely linked to the biology and physiology of any type of biological system utilising the transmission of spikes or action potentials as a form of communication between two points. The surrounding functional blocks were then designed to take advantage of channel characteristics and to reduce the effects of unreliable transmission encountered between successive neurons.

First, the channel model is defined after which the definition of the noise model follows.

5.3.1 Nature of Channel Model

The channel model can be derived by considering the physiology of the auditory nerve bundle.

5.3.1.1 Physiology of auditory nerve bundle

The auditory nerve bundle can be divided into *afferent* and *efferent* fibers, where in the case of the afferent fibers information is carried from the organ of Corti to the brainstem and brain, and in the case of efferent fibers information is brought from higher neural centres to the periphery (Yost, 2000).

There are approximately 12 000 outer and 3 000 inner haircells (Spoendlin, 1974). 90 to 95 % of the afferent fibers innervate the inner haircells (Spoendlin, 1974) indicating that the inner haircells carry information from the cochlea to the cochlear nucleus. It is however not clear what role the outer haircells play but it has been observed by Moody, Stebbins and Hawkins (1976) that damage of the outer haircells causes a significant loss in sensitivity and frequency resolution. More recently effects of outer haircell impairment on auditory nerve responses have been documented by (Liberman, 1984; Liberman and Dodds, 1984a; Liberman and Dodds, 1984b; Liberman and Kiang, 1984).

5.3.1.2 Redundancy

Human inner haircells synapse onto between 6 to 8 auditory nerve fibers (Nadol, 1983) and cat inner haircells synapse onto up to 20 auditory never fibers (Liberman, 1982). This suggests that a form of redundancy is present at this stage of the auditory system. Another form of redundancy is provided by population codes, which were discussed in *Chapter 3*. To simplify matters it is assumed that one haircell is innervated by *only* one nerve fiber. This reduces the computational load placed on the model.

5.3.1.3 Spike rates

Each cell needs a certain time to recover between firings. The delay between successive spikes is called the absolute refractory period or “dead-time” in which no other action potentials or spikes can occur. The relative refractory period follows the absolute refractory period in

which the nerve fiber is yet again sensitive to input stimuli but only to strong ones. Measured spike trains observed by Johnson (1996) suggested a refractory period of 4 ms and a maximum spike rate of 250 spikes per second. It has also been shown that theoretical maximum spike rates of 300 to 1 000 spikes per second are possible given the maximum refractory periods of 3 ms to 1 ms (Flanagan, 1972; Rieke et al., 1997).

5.3.2 Noise Model

The noise model is one of the most complex models in this dissertation. The reason for this is because of the daunting task of making sense of neural coding by observing how neurons in the CNS code signal information and to distinguish the signal from the noise by observing the firing patterns (Lestienne, 2001). As was already discussed in *Chapter 3* it is not quite clear what neural code is present in the CANS or CNS and unless a neural code can clearly be found the boundary between signal and noise is vague.

In fact it was shown by Cecchi, Sigman, Alonso, Martinez, Chialvo and Magnasco (2000) that noise is dependent on the type of message signal being coded by a neuron, where a leaky integrate-and-fire (LIF) neuron (see *Subsection 5.4.1*) was used with a noisy threshold voltage. The reliability of the neural coding mechanism was tested where the reliability is defined as the ability to produce the same firing pattern when presented with the same stimulus. While temporal patterns were unreliable when the injection current was constant, they became more reliable with a random input current. Similar simulations by Tiesinga, Fellous and Sejnowski (2002) showed that reliability was indeed dependent on the message signal.

The dependence of the variability of spike trains, given a particular message signal is however not solely due to the noisiness of the neuron but can be attributed also to the fact that neurons are tuned to certain types of inputs and not to others such as the phenomenon of tuning curves in the CANS suggests (Yost, 2000). Although new methods are being found (Schreiber, Fellous, Whimer, Tiesinga and Sejnowski, 2003; Alstrøm, Beierholm, Nielsen, Ryge and Kiehn, 2002) to measure the reliability of neurons, they only underline the dependency of spike generation on a message signal. Thus the source of "noise" requires careful scrutiny as is done next.

5.3.2.1 "Noisy" spike trains

The individual neuron's action potential has an all-or-none characteristic, which means that the amplitude of the neural discharge does not vary with the level of the stimulus. The amplitude of the action potential magnitude is typically 100 mV in the squid (Hodgkin and Huxley, 1952). The regeneration of action potentials at the axons continually compensate for losses in action potential magnitude so that the action potential suffers no decreases in amplitude (Saito and Noguchi, 1981; Rasminsky, 1978) during transmission over myelinated and unmyelinated networks of fibers. Changes in the shape and size of action potentials are thus not the source of noise but rather the *intracellular* noise and *extracellular* noise, where the former characterises the noisy process of spike generation such as temperature and ion fluctuations. On the other hand, the latter describes the reliability of the transmission of the action potential across a "channel", which in this case is across the axon and dendrites of a neuron.

Intracellular noise

Intracellular noise is generally characterised by noisy thresholds, noisy reset positions, a noisy integration process and synaptic transmission failure, where the last item signifies an all or none loss of a spike (Verveen and Derksen, 1968; Verveen and DeFelice, 1974).

It has been shown by Azouz and Gray (1999) and Azouz and Gray (2000) that the membrane potential fluctuations spans a range of 15-25 mV in cortical neurons but that these fluctuations are remarkably synchronised among neighbouring cells (Lampl, Reichova and Fenster, 1999), which could hint to higher underlying statistics rather than just randomness or noise.

Extracellular noise

Extracellular noise is generally characterised by spike-time spreading. Spike-time spreading occurs between successive neurons and is a function of retransmission delays and conduction velocities given a particular type of axon or dendrite. These delays have shown to be in the microsecond range in transcallosal fibers by Houzel, Milleret and Innocenti (1994) indicating a highly reliable conduction process from one synapse to another.

Much research has been done into measuring spike trains and determining the amount of stochasticity present relative to the deterministic component to ultimately determine what neural code might be implemented by various neurological communication channels (Kohn, 1997). Some have suggested that only rate information is coded by spike trains and when corrupted by noise, results in a Poisson point process (Johnson, 1996; Snyder, 1975). In more recent experiments with retinal ganglion cells of rabbits and salamanders, it was shown that the spike trains could not be entirely characterised by their spike rates (Lestienne, 2001; Rieke et al., 1997; Deco and Schürmann, 1999) and that an underlying temporal correlation was present. This meant that the "noise" as described by Poisson distributed point processes was indeed not entirely noise but hinted at an underlying higher-order encoding process. Determining the nature of stochasticity of a measured spike train is thus no trivial task and may differ for different parts of the CNS.

Gaussian noise

The location of where noise is introduced during the neural coding and transmission process was defined in the previous section by identifying both intracellular and extra cellular noise. The statistics of the noise are however chosen to be of a Gaussian nature for all types of noise present. This simplifies the noise model and avoids designing more non-white shaping filters for the estimator in *Chapter 6*. The simplification made are similar to the ones implemented by Bialek et al. (1991), Cecchi et al. (2000) and Tiesinga et al. (2002). Bialek et al. for example introduced Gaussian jitter noise to the spike train with a standard deviation of 2 ms where the maximum spike rate of the neuron lay between 100 and 200 spikes per second. Cecchi et al. and Tiesinga et al. on the other hand introduced Gaussian noise during the spike generation process by a LIF neuron, simulating a noisy threshold.

The limitations of the channel model are summarised next.

5.3.3 Channel Model Characteristics

Given the channel and noise models in *Subsection 5.3.1* and *Subsection 5.3.2* the limitations can be summarised as follows:

1. An action potential or spike is represented by a Dirac-delta function,
2. where the maximum spike rate of a single neuron is 500 spikes per second.
3. Both the absolute and relative refractory periods are equal to 1 ms (2 ms in total),
4. the type of noise is additive white Gaussian noise (AWGN),
5. the noise found on the different channels is statistically independent from one another and
6. spike transmissions are unreliable and spikes are lost at random.

5.4 ENCODER, DECODER MODEL

5.4.1 Leaky Integrate-and-Fire Model

The leaky integrate-and-fire neuron model (LIF) is one of the more simple neuron models found in the literature (Eliasmith and Anderson, 2003; Maass and Bishop, 1999) and is used to encode a time-dependent variable. The neuron model takes into consideration the capacitance between the fluids at the inside and the outside of the cell membrane and both the capacitive and leakage currents. The leakage current flows through a leakage resistor, modelling the effect of ion channels in the cell membrane, where sodium, potassium and chloride ions migrate through.

In an electrical sense the neuron model can be though of as a current source driving a parallel resistor/capacitor combination as shown in *Figure 5.2*.

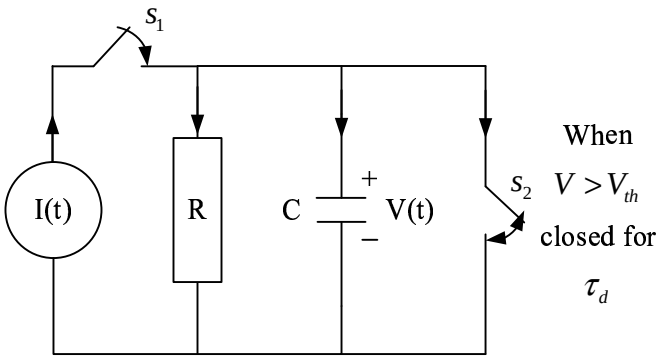

Figure 5.2: Leaky integrate-and-fire model.

A mathematical interpretation of *Figure 5.2* is

$$RC\frac{dV(t)}{dt} + V(t) = RI(t). \tag{5.1}$$

As the current gradually charges the capacitor the voltage across it, representing the action potential of the incoming spike train, rises exponentially until reaching a voltage of $I(\infty)R$. The time required to reach a steady-state depends on the time constant τ_{RC}. In order to define a spike, a threshold voltage V_{th} needs to be defined, which the rising voltage $V(t)$ reaches at time t_f, signifying a spike time. After the neuron has fired, generating an action potential or spike at the output, switch s_2 closes for a refractory period τ_d inhibiting additional firings. In this case the value to which the voltage $V(t)$ is reset is taken to be zero. The action potential generated by the LIF neuron can be defined independently from the circuit in *Figure 5.1* since it only determines the incident at which a spike is created and not the shape of the spike. In this case only the spike times are relevant (see *Section 5.5*).

5.4.2 Optimal Linear Estimators

After having chosen a particular coding strategy, the next step would be to see how the spike trains can be utilised to reconstruct the original continuous-time signal. This enables one to evaluate for instance how well a particular code performs in a noisy environment. Although signal reconstruction may or may not be important for the brain, since so little is known about how the brain uses the information coded by the spikes, it provides a tool to understand its robustness to errors and the information rate at which it conveys its message.

The first question that comes to mind is: In what way can a continuous-time signal, encoded by discrete spike times possibly be decoded to obtain the original signal? A form of interpolation is required, that predicts what the signal might have been, regarding a particular spike train. This clearly requires a probabilistic approach to the problem in that one wants to find the probability of a particular signal $s(t)$, given a particular sequence of spike times $\{t_i\}$ or

$$\mathbf{P}[s(t)|t_i]. \tag{5.2}$$

In order of the probability of *Equation (5.2)* to be of any use, the signal has to be uniquely defined by a particular spike train, resulting in a clearly distinguishable peak of the conditional distribution of **P**. Any ambiguities might result in finding the incorrect waveform $s(t)$. Using maximum likelihood estimation (MLE) or finding the minimum mean-squared error (MMSE), provides tools to estimate the original signal.

5.4.3 MMSE

For the MMSE estimator it is desirable to estimate the signal $x\hat{(}t)$ by observing the spike train

$$p(t) = \sum_{n=0}^{N-1} \delta(t - t_n). \tag{5.3}$$

This requires that for the linear reconstruction

$$\hat{x}(t) = \int K_1(\tau) \sum_{n=0}^{N-1} \delta(t - \tau - t_n), \tag{5.4}$$

to find the kernel K_1 that provides the best estimate in the MMSE sense. *Equation (5.4)* can be simplified to

$$\hat{x}(t) = \sum_{n=0}^{N-1} K_1(t - t_n). \tag{5.5}$$

The kernel $K_n(t)$ specifies not only the nature of the neuron but also the nature of the signal and its surroundings. The dependency of the kernel on time, allows for an adaptation in the environment, but requires continuous updating of the kernel. *Equation (5.5)* however is a *linear* reconstruction filter where a non-linear version would be of the form

$$\hat{x}(t) = \sum_{i} K_1(t - t_i) + \frac{1}{2} \sum_{ij} K_2(t - t_i, t - t_j) + \dots. \tag{5.6}$$

Higher-order terms can be modelled by $K_2, K_3, \dots, K_n$. A method for finding the best estimation of $\hat{x}(t)$ by observing a spike train was developed by Kolmogoroff and Wiener (Rieke

et al., 1997). In particular if the signal and noise are of Gaussian nature and the neuron has a linear input/output relationship then an optimal linear estimator exists where the first kernel is found to be

$$K_1 = \frac{\frac{E\left[x^2\right]}{E\left[n^2\right]}}{\frac{E\left[x^2\right]}{E\left[n^2\right]} + 1}, \tag{5.7}$$

where n is the noise signal and

$$SNR = \frac{E\left[x^2\right]}{E\left[n^2\right]}. \tag{5.8}$$

As soon as the information signal, the noise signal or both are not of Gaussian nature, optimal linear prediction is not possible anymore, even if the model for the neuron is a linear one. The statistics of the signal and its environment is thus of utmost importance when designing a decoder. Knowledge of the exact intrinsical workings of the encoder are thus not necessarily a prerequisite as was argued by Eliasmith and Anderson (2003) and Rieke et al. (1997), to finding a suitable optimal linear predictor.

5.4.4 Correlation Times

The decoding strategies also depend on the correlation time. If the neural response produces a dense spike train and the number of spikes relative to the correlation time is larger than one, then the extrapolated knowledge surrounding a spike overlaps with other spikes, making reconstruction not feasible. On the other hand, if the correlation time is very small then a narrow reconstruction filter as seen in the time domain will be more prone to "jitter" noise.

The reconstruction filter for the case where the correlation time is small relative to the ISI, can be found by the reverse correlated function, while it becomes a bad approximation if the correlation time becomes too large. There thus exists a tradeoff between the two. From a filter design perspective, a wide-band LPF found by using a very narrow impulse response will allow for high frequency estimation. If the higher frequency components were however predominantly noise, a narrower LPF would remove these components. The organism is however pressed for time thus a relatively fast decision time is desirable, which would make it

prone to noise.

There is thus no ideal optimal linear predictor for all cases but it has been shown by Eliasmith and Anderson (2003) that a filter for a shorter correlation time also works well for longer correlation times. These results are however based on the assumption that no noise was present on the spikes themselves. The optimal filter was also found to be non-causal, something that is not biologically plausible. A spike is only allowed to influence the estimate once it has occurred while the stimulus influences the generation of the spike before the spike occurs. Causality however introduces *delays* in the estimation, something that has to be kept in mind when evaluating the estimator.

Generally the coder model together with the decoder model are limited by the channel model, which was defined in *Section 5.3*. In this case the channel is limiting the amount of information that can be transmitted at any given time due to noise added to the system. The upper bound of the channel capacity can be obtained by viewing the channel as a generic digital communication system (Proakis, 1995; Haykin, 1994). It has been shown that by using intricate coding schemes to encode the stream of information, a higher bandwidth efficiency can be obtained. In the next few sections the type of discrete time to spike-time coding that was implemented is discussed but first a look will be taken at issues related to the coder and optimal estimators for spike trains.

5.4.5 Linear Estimators

Finding an optimal linear estimator has been the focus of many a research and has been extensively documented by amongst others Rieke et al. (1997) and Eliasmith and Anderson (2003). Optimality however suggests having found the best solution given a problem and its set of limitations. The way performance is classified and measured shapes the results and different methods result in different solutions. It is thus of utmost importance when designing optimal estimators for spike trains, that certain criteria's are adhered to, enabling a comparisons to results obtained previously by others.

The information theoretical approach by Shannon and Weaver (1949), provides a quantitative measure of how much information was present at a receiver Y, given an input X. Although the derivations by Rieke et al. (1997), Eliasmith and Anderson (2003) are made,

given a Gaussian source and channel, these methods provide an *upper* bound to the amount of information that can possibly be transmitted by means of a particular spike train. In this chapter only a short look will be taken at optimal estimators, since they are required when designing postsynaptic current (PSC) filters. The aim here however is not to find an optimal estimator, but rather a physiological plausible estimator. Nonetheless, the optimal estimator provides a useful comparison to the suboptimal PSC filter.

The derivation of the optimal estimator was included solely for the sake of interest and can be found in *Addendum A.3* with relevant references.

5.4.5.1 Postsynaptic current filter

The PSC filter, unlike the optimal filter derived by Eliasmith and Anderson (2003), is a much more plausible explanation of how the physiology would decode a spike train. For one, the optimal filter is a non-causal filter suggesting that decoding of a spike occurs before an onset of a spike, something that is highly unlikely. The PSC on the other hand is a causal filter. There are various waveforms that resemble the exponentially decaying PSC such as the exponential kernel (Abbott, 1994)

$$K(t) = \frac{1}{\tau}\, e^{-t/\tau}, \tag{5.9}$$

and alpha-function kernel

$$K(t) = \frac{t}{\tau^2}\, e^{-t/\tau}. \tag{5.10}$$

While the exponential kernel is the least complicated one of them, having only one variable, it produces a discontinuity at the beginning, which is eliminated by the alpha-function kernel. Unlike the optimal filter, the PSC filter has a very specific magnitude response, which is not optimally chosen for each particular frequency. Thus an inferior PSC filter, in the MMSE sense, would be expected, as compared to the optimal filter. Also since the magnitude response is fixed to a particular value for τ an optimal τ for a particular ensemble of input signals would be expected.

Noisy spike trains in the context of the LIF neuron coding

As was discussed in chapter *Section 5.3*, two types of noise models were implemented, one introducing "jitter" on each individual spike and another randomly dropping spikes. Noise is an important aspect for spike train decoding, which needs to be considered when designing a decoding filter. Since the optimal and PSC filters act as lowpass filters it is expected that high frequency noise is successfully attenuated. However, the smaller the membrane time-constant the larger the passband of the linear estimators becomes, which increases the noise sensitive of the decoder. It is thus imperative to study the performance of a decoding filter under all conditions in particular to investigate the effects of

1. spike-time "jitter",
2. random discarding of spikes and
3. linear decoding errors due to a non-linear encoding process.

5.4.5.2 Coding the signal

The speech signal was decomposed into separate but overlapping spectral bands (see *Chapter 4*). In this dissertation the interest only lay in the amount of power found in a particular band, which was obtained by demodulating the entire frequency content within a particular band to a single DC component. Two variables need to be coded, the time-varying DC component and the frequency at which the particular component is found. The frequency dependent spatial representation (tonotopic organisation) implicitly "conveyed" the frequency dependency of each spectral band. It was assumed that an external higher-level system inherently knows of such a tonotopic organisation. Such an assumption is plausible since in the auditory system at frequencies above 1 kHz (Rose, Brugge, Anderson and Hind, 1967), phase-locking diminishes rapidly as the frequency increases and intensity coding becomes the dominant information being transmitted via the spike trains. The time-varying DC component was coded by a LIF neuron, resulting in a time-varying spike rate.

Even though only a single variable, the time-varying DC component, is coded by a time-varying spike rate does not imply that specifically a *rate* code is begin used as defined in *Chapter 3*. The reason for this is simple. For the LIF neuron (see *Equation (5.1)*) $I(t)$ is the time-dependent input variable. As was discussed in *Subsection 4.4.1* for any periodic signal

there exists a Fourier series and the Fourier series is a sum of sinusoids. A sinusoid can be fully described by its amplitude, frequency and phase. Thus by encoding a sinusoid with the LIF neuron by using *Equation (5.1)*, three or more variables can be coded simultaneously in a spike rate. The *change* in spike rate, as in the single variable case, carries additional information about the signal, which when decoding takes place, can be extracted for signal reconstruction purposes.

However, does the way the signal is coded, in this case using a LIF neuron to code a time-varying DC component, not predetermine the type of code used? To identify the ambiguity it will be argued for both a temporal code and a rate code.

In the case of a rate code, it can be argued that since only one variable is being coded and the rate of spikes is proportional to the magnitude of the input current, the code being used is a rate code. By averaging over a time window the spike rate can be extracted and noise removed from the spike train. This decoder, a PSC filter can be thought of as an averaging tool, where τ determines the width of the averaging window. But does the precise timing of the spikes play any significant role here? Yes, if a sinusoidal input signal was coded by means of a time-dependent spike rate, the spike rate would vary accordingly, where the position of its maxima can be used to indicate both amplitude and frequency of the sinusoid. By averaging over a window larger than the period of the signal, the identity of the sinusoid would be lost. This suggests that spike-time intervals are of great importance, something that is associated with a time code. Thus the neural code that is used here has both the characteristics of a *rate* code and a *temporal* code. Thus by having refrained from choosing a particular code in *Chapter 3* a code having both rate and temporal characteristics has been discovered.

5.5 SIMULATION

5.5.1 Introduction

In this section the reliability of neural coding implemented by a LIF neuron is evaluated in terms of intracellular and extracellular noise. However, before this can be done a suitable PSC filter has to be found. This is done by finding an optimal filter first from which intuitively the design of the PSC filter follows. The focus in this case is on the biologically plausible PSC filter since optimal decoding has already been investigated extensively by others (Rieke et al., 1997; Eliasmith and Anderson, 2003).

The noise added to the spikes combines the effects of the noise sources discussed in *Subsection 5.3.2*. Although it was mentioned that intracellular noise is dependent on the type of message signal applied only its *effect* on the spike train is implemented here. The "jitter" added to the spike times combines the effect of a noisy threshold, reset position and integration process all of whom are intracellular noise sources. The spike time spreading, an extracellular noise source can also be included here, since spreading of the individual spikes can be logically implemented by adding "jitter". Synaptic transmission failures (intracellular noise) however cannot be simulated as a form of spike-time jitter. Instead an additional method of dropping spikes is added to account for this type of noise.

5.5.1.1 Dropping of spikes

The dropping of spikes investigates the effect that a loss of spikes from the original spike train has on the reliability of the signal being transmitted. Spikes are dropped by a uniformly distributed all-or-none random process where each spike is assigned a random variable lying between zero and one. A cutoff threshold is then picked between zero and one according to the desired number of spikes that need to be dropped. The dropping of spikes simulates the synaptic transmission failure that was discussed earlier.

5.5.1.2 Spike-time "jitter"

In order to investigate the behaviour of an error in the exact occurrence of spike times, each individual spike was displaced by a random amount obtained from a Gaussian distribution

with a fixed variance. The variance was chosen in such a way that the standard deviation of the error was at most slightly larger than the smallest spike interval. No spikes were however dropped if the position of two adjacent spikes were reversed in the process of generating the position errors. This process can be seen as simulating the noisy threshold, noisy reset positions and noisy integration process that was discussed earlier.

5.5.2 Leaky Integrate-and-Fire Model Implementation

5.5.2.1 Coding

In *Addendum A.4* the behaviour of the LIF is derived and can be represented by the convolution integral

$$V(t) = \frac{R}{\tau_{RC}} \int_0^{T_f} e^{-(t-t')/\tau_{RC}} x(t') \, dt'. \tag{5.11}$$

In order to apply the convolution integral in the discrete-time domain, a discrete version of *Equation (5.11)* is required, which can be realized in the form of a *tapped-delay-line filter* implementation. Defining $t = nT_s$ and $t' = kT_s$ where the filter impulse response (T_f) is assumed to be of finite length $T_f = NT_s$.

The discrete-time form is given by

$$V(nT_s) = \frac{1}{C} \sum_{k=0}^{N-1} e^{-kT_s/\tau_{RC}} x(nT_s - kT_s) T_s. \tag{5.12}$$

The spike times are obtained from *Equation (5.12)* by observing when $V(nT_s)$ reaches the threshold voltage, discussed in *Section 5.4*. The exact position in time in which the membrane voltage reaches the threshold voltage is obtained by means of interpolation of which the resolution is dependent on the temporal resolution of the integral. The resulting spike times can be thought of as point processes at positions in time given by

$$p(t) = \sum_{n=0}^{N-1} \delta(t - t_n), \tag{5.13}$$

where t_n represents the spike positions in time.

5.5.2.2 *Suitable parameters*

The values for the time constant τ_{RC} can be found in Deutsch and Deutsch (1993) and are given in *Table (5.1)*.

Table 5.1: Typical neuron model parameters (unmyelinated).

Component	Value
$MaxRate$	500 Spikes/s
V_{th}	90 mV
R	4 kΩ
C	1 μF
τ_{RC}	4 ms

Care however has to be taken when using these values since for a maximum spike rate of 500 spikes per second a constant current of 57 μA is required. If the current is larger, the spike rate will increase. The current source driving the resistor/capacitor combination can thus be though of as being the voltage as found at the output of the lowpass filter (*Section 4.5*) (measured in volts and limited to a maximum of 1 V peak) divided by an appropriate resistor to limit the operation of the LIF neuron to the active region as will be shown in *Figure 5.3*.

5.5.2.3 *Neuron pairs*

One of the limitations of the LIF neuron implemented by *Equation (5.11)* is that the voltage $V(t)$ only rises above a certain threshold V_{th} if the input $x(t)$ is positive. Similarly if the threshold voltage and exponential function in *Equation (5.11)* would undergo a sign reversal, then the voltage $V(t)$ would only drop below the threshold voltage if the input signal were to be negative. Any signal that has both negative and positive components thus requires two LIF neurons to code both positive and negative values. It is however not unusual to find "on" and "off" neurons encoding the same stimuli as was discussed by Eliasmith and Anderson (2003), Bialek et al. (1991) and Rieke et al. (1997). Thus the total range over which the neuron pair is active is now doubled and a centre of reference becomes less significant.

In the outer ear, different types of spontaneous rate auditory nerve fibers are grouped together to code the mechanical to neural transduction of a single inner haircell. In the case of a low frequency stimulus, which can drop below zero, the drive is simply lost by the low spontaneous rate fibers. However, in the case of high spontaneous rate fibers, which have a "DC bias" by virtue of their spontaneous firings, a negative input stimulus is coded as a *drop* in discharge. However, at a conceptual level the exact biological method of mechanical to neural transduction and the subsequent neural coding becomes less important since neuron pairing or grouping of different spontaneous rate auditory nerve fibers only serves to increase the dynamic range of the input stimulus.

Consequently, in this dissertation it was assumed that two neurons encode the positive and negative parts of the signal separately. Even though the quantity to be coded (see *Subsection 4.5.1*), namely the amount of power within a spectral band, can intuitively never be negative since the signal is squared mathematically, the impulse response of the LPF following the demodulator does go negative. In the absence of positive power within the band, the neurons will fail to fire unless a second neuron codes the negative power as well.

5.5.2.4 DC level offset

Using pairs of neurons coding signals having both negative and positive amplitudes is one way of solving the problem. Another is to simply work with positive signals. Any signal that does contain negative components is simply translated by a DC offset.

5.5.2.5 Dynamic range

From *Addendum A.4*, *Equation (A.28)* it is known that for a *constant* current I_0

$$V(t) = RI_0\left[1 - e^{-(t\,-\,t_0)/\tau_{RC}}\right], \tag{5.14}$$

where I_0 represents the ionic current through the membrane and $V(t)$ the membrane voltage across the membrane surrounding a neuron (Deutsch and Deutsch, 1993).

From *Equation (5.14)* the "on" neuron will thus only fire if

$$I_{min} > \frac{V_{th}}{R}, \tag{5.15}$$

and a constant current of

$$I_{max} = \frac{V_{th}}{R(1 - e^{-1/MaxRate \cdot \tau_{RC}})}, \tag{5.16}$$

is required to produce the maximum spike rate $MaxRate$. The current driving the "on" neuron thus has to be in the range

$$\frac{V_{th}}{R} \leq I_0 \leq \frac{V_{th}}{R(1 - e^{-1/MaxRate \cdot \tau_{RC}})}. \tag{5.17}$$

It is clear from *Equation (5.17)* that only the membrane time constant τ_{RC} and the maximum spike rate determine the *dynamic range* of the coder. It is however evident from *Equation (5.17)* that when the current is below V_{th}/R the membrane voltage $V(t)$ will never reach the threshold voltage to produce a spike thus operating in the inactive region as shown in *Figure 5.3*.

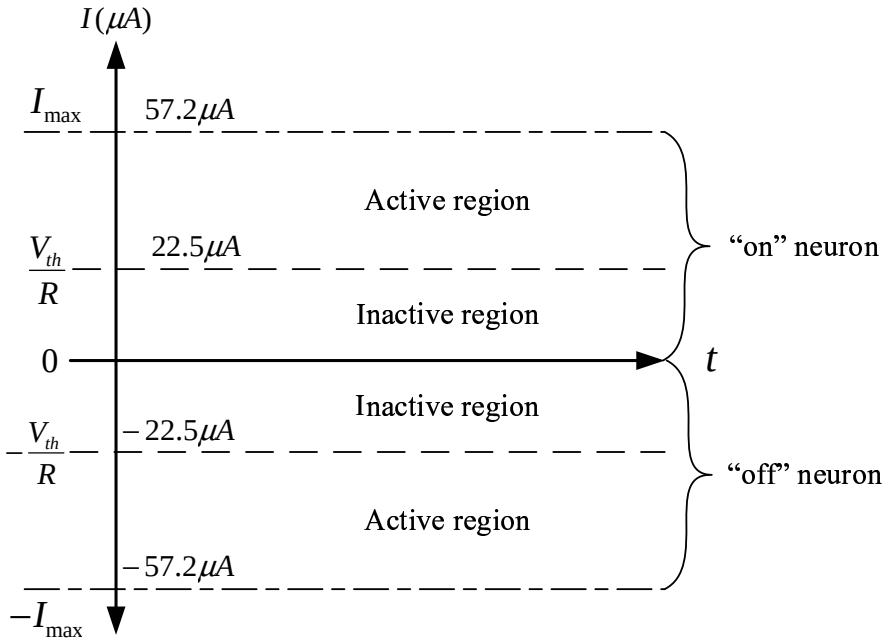

Figure 5.3: Dynamic range of "on/off" LIF neuron pair without bias current.

A slight adaptation of *Equation (5.14)* is required to ensure that any current waveform I_0 over the entire range, I_{min} to I_{max} will produce a spike, which requires an input current in the active region of *Figure 5.3*. This is achieved by adding a bias current I_{bias} so that

$$I_{min} + I_{bias} \geq \frac{V_{th}}{R}. \tag{5.18}$$

I_{min} is in this case zero, since an identical pair of neurons is used, though with opposite signs. I_{min} on the other hand represents the boundary between the two neurons, which is chosen at zero. *Equation (5.14)* then becomes

$$V(t) = R\left[\mathbf{g} \cdot I_0 + I_{bias}\right]\left[1 - e^{-(t\,-\,t_0)/\tau_{RC}}\right]. \tag{5.19}$$

Equation (5.19) contains a bias term that moves the operation of the LIF neuron relative to the input current $I(t)$ into the active region. The gain ($\mathbf{g}$) ensures that the combined current of I_0 and I_{bias} never exceeds some maximum value over which the spike rate stays constant (region above active region in *Figure 5.3*). It is however evident that the non-linear region of operation has now changed and has to be taken into consideration if a *non-linear* estimator were to be found.

Given the parameters found in the physiology (from *Table (5.1)*), the range in which the current can lie according to *Equation (5.14)* and *Equation (5.19)* is summarised in *Table (5.2)*.

Table 5.2: Dynamic range of I_0 in a single LIF neuron ($\mathbf{g} = 0.61$).

Equation	μ**A**	μ**A**	**dB**
$V(t) = RI_0\left[1 - e^{-(t\,-\,t_0)/\tau_{RC}}\right]$	22.5	57.2	8.1
$V(t) = R\left[\mathbf{g} \cdot I_0 + I_{bias}\right]\left[1 - e^{-(t\,-\,t_0)/\tau_{RC}}\right]$	0	57.2	35.2

When a neuron pair is used, more than double the current range can be attained (35.2 dB instead of 8.1 dB).

The visual difference between the two equations in *Table (5.2)* is shown in *Figure 5.4*.

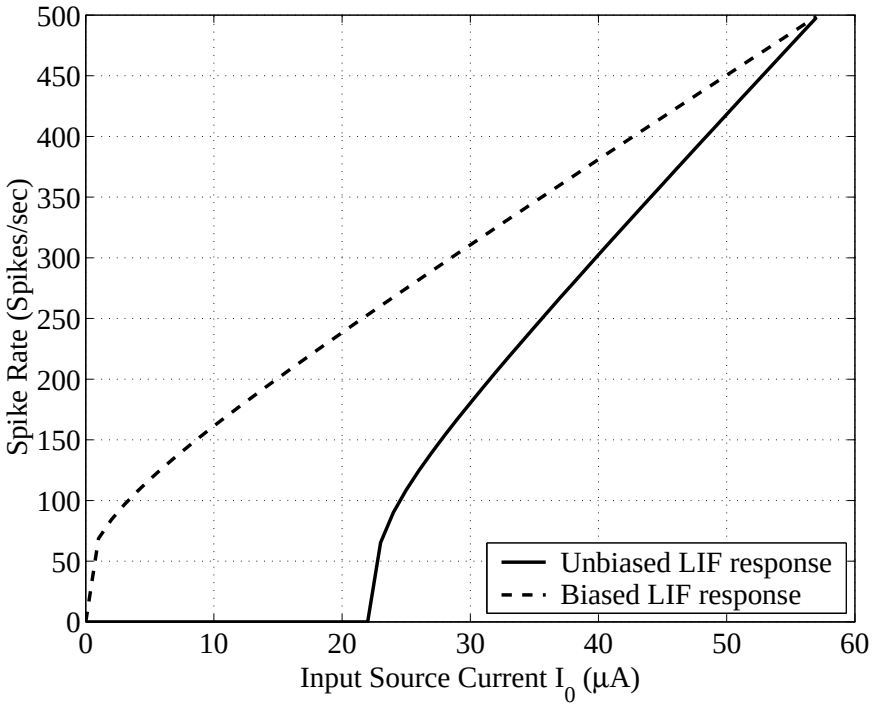

Figure 5.4: Comparison between biased and unbiased LIF neuron equation.

Both biased and unbiased waveforms depict the non-linear relationship that exists between the input current I_0 and the spike rate. As expected, the biased LIF neuron fires at any input current between 0 and $57.2\mu A$ and thus operates in the active region of *Figure 5.3* whereas the unbiased waveform requires an input current of between 22.5 and $57.2\mu A$ in order to fire as noted in *Table (5.2)*. The bias however does not change the form of the input/output relationship and merely "stretches" it across a larger range of input currents. The nature of the LIF neuron's response to an input current shown in *Figure 5.4* also compares favourably with measurements made on real neuron (Johnson, 1996).

5.5.3 Decoding

In order to obtain an estimate of the spike train coded original signal $x(t)$, a filter $h(t)$ is required such that

$$\hat{x}(t) = \int_0^T h(t-t') \sum_n \delta(t'-t_n)\, dt'. \tag{5.20}$$

While it has been shown by Eliasmith and Anderson (2003) that an optimal filter can be found (see *Addendum A.3* for derivation), it remains a complicated task requiring a Monte Carlo analysis of a large set of possible inputs. The optimal filter is non-causal and thus not biologically plausible. It has however also been shown by Eliasmith and Anderson (2003) that a linear filter related to the postsynaptic currents provides similar results in terms of decodable information (only 6% less than optimal filter) and RMS error (about twice that of optimal filter). The PSC model is given by

$$h_{PSC}(t) = \frac{1}{\tau}\, e^{-t/\tau_{syn}}, \tag{5.21}$$

where τ_{syn} is the synaptic time constant. In fact, to find a linear filter it is only required that kernel K satisfies

$$\int_0^\infty K(t)\, dt = 1. \tag{5.22}$$

From *Equation (5.22)* there are infinitely many kernels that satisfy this equation.

It is however interesting to note that *scaled versions* of the exponential and alpha-function kernel, mentioned earlier and the difference of exponentials kernel

$$K(t) = \frac{1}{\tau_1 - \tau_2}[e^{-t/\tau_1} - e^{-t/\tau_2}]. \tag{5.23}$$

satisfy this condition. The exponential kernel that was used by Eliasmith and Anderson however has one drawback. It contains an instantaneous jump at $t = 0$ and thus a discontinuity, which is eliminated by the two other kernels.

In this case only the potential of the first two kernels is explored as an alternative decoding filter to that of the optimal linear filter. The first two kernels, PSC1 and PSC2 can be altered by either changing the time constant or the amplitude of the impulse response resulting in an adjustable gain in the frequency domain. No other possibilities exist to match the PSC filters with the optimal linear filter.

In *Equation (5.9)* and *Equation (5.10)* the areas of the functions are normalised to one since this normalises the gain in the frequency response seeing that

$$\int_{-\infty}^{\infty} g(t)dt = G(0), \tag{5.24}$$

where $G(f)$ represents the Fourier transform of equation $g(t)$. The normalised PSC filter characteristics are summarised in *Table (5.3)* (see *Addendum A.5* for derivation).

Table 5.3: Frequency and area characteristics of PSC filters.

PSC type	Fourier transform	Area
Exponential (PSC1)	$\frac{1}{1+j2\pi f\tau_{RC}}$	1
Alpha-function (PSC2)	$\frac{1}{(1+j2\pi f\tau_{RC})^2}$	1

From this point onwards the exponential kernel will be referred to as the PSC1 filer and the alpha-function kernel filter as the PSC2 filter. To observe how the PSC1 filter characteristics change with changing time constant, both the time and frequency response are shown in *Figure 5.5*.

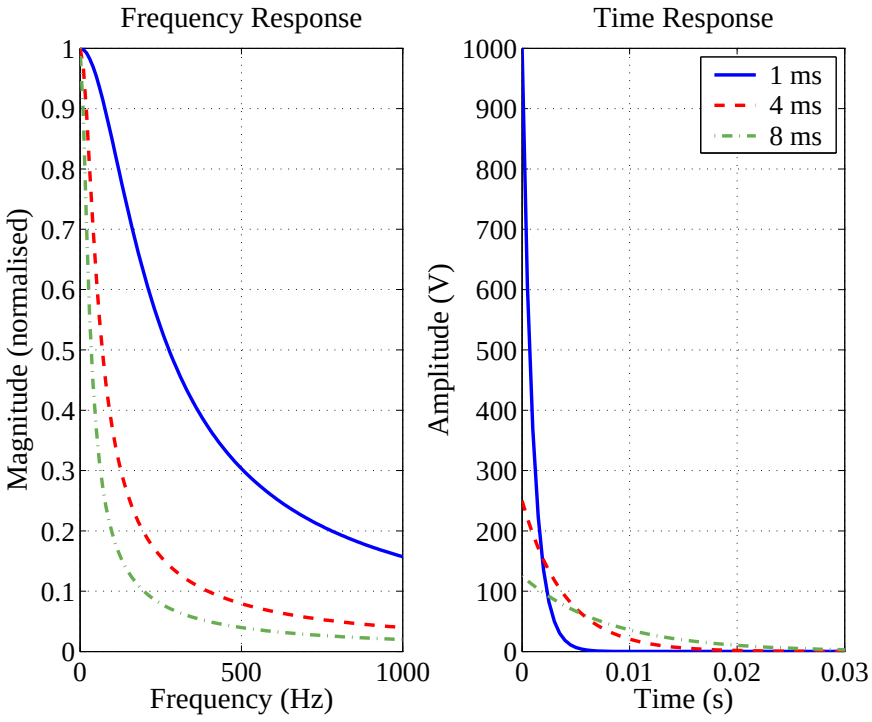

Figure 5.5: Time and frequency response of PSC1 filter for different time constants.

Choosing a smaller time constant results in a wider passband and a faster decaying impulse response. Similarly by choosing a larger time constant, the passband is decreased and the impulse takes longer to decay to zero.

The effect that the time constant has on the phase response of the filter is depicted in *Figure 5.6*.

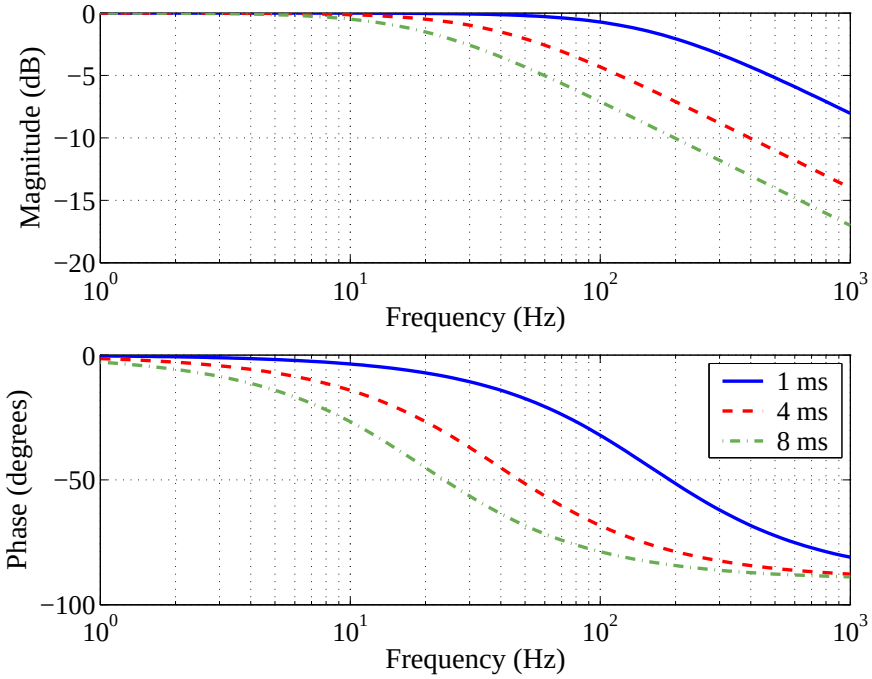

Figure 5.6: Frequency and phase response for PSC1 filter.

The PSC1 filter has a non-linear phase response where the maximum slope of the phase increases with an increasing time constant.

Similarly, to observe how the PSC2 filter characteristics change with changing time constant, both the time and frequency response are shown in *Figure 5.7*.

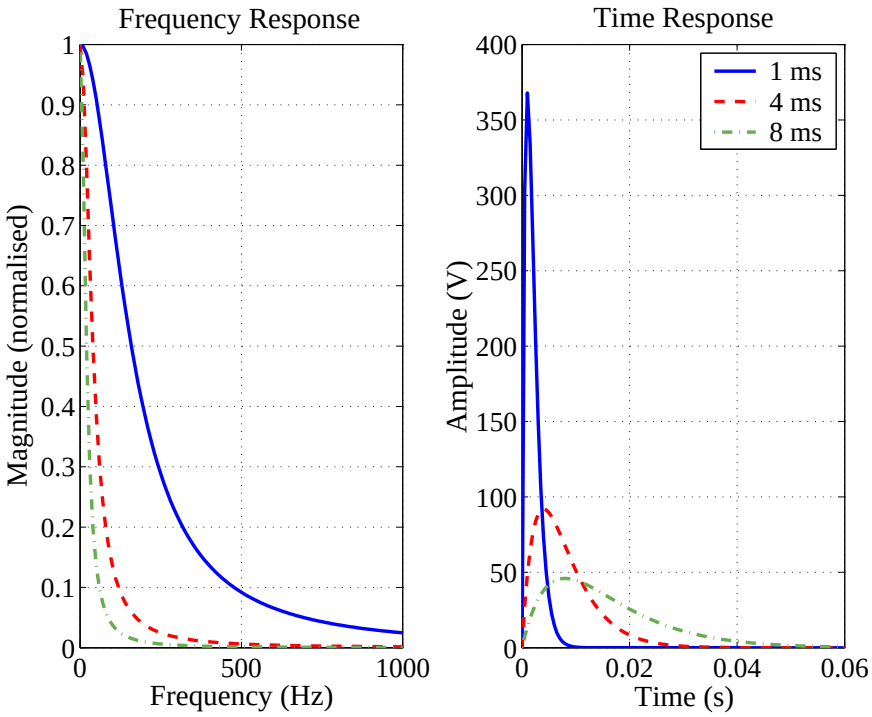

Figure 5.7: Time and frequency response of PSC2 filter for different time constants.

In case of the second PSC filter the roll-off is much steeper and increases with increasing time constant. Unlike the first PSC filter the impulse response of the second filter has a finite rise time. Choosing a smaller time constant results in a faster rise and decay time.

To observe the effect that the time constant has on the phase response of the filter the frequency and phase response of the filter are depicted in *Figure 5.8*.

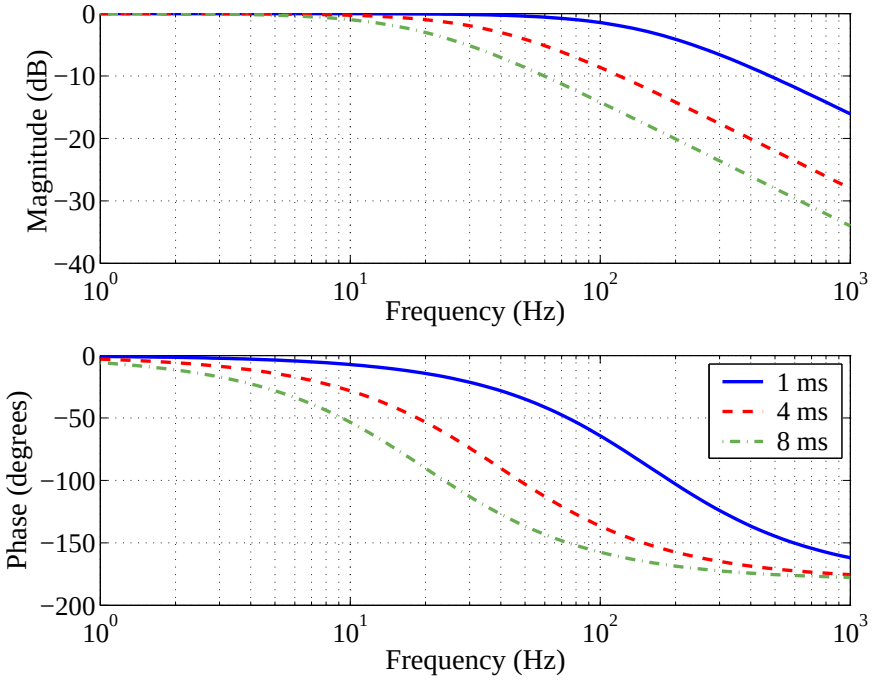

Figure 5.8: Frequency and phase response for PSC2 filter.

Similarly to the phase response of the first PSC filter, the phase response is non-linear but the phase increases much faster with an increasing frequency. Comparing the two PSC filters on grounds of obtained results, it can be seen that the first PSC filter has a wider passband than the second PSC filter but introduces less phase distortion for the same time constant. The first PSC filter, however, due to its instantaneous change in amplitude, is physically unrealisable and thus biologically implausible. Both PSC filters are however used as a mean of comparing the performance of the two.

5.5.3.1 Scaling PSC decoding filters

PSCs however do not occur as normalised waveforms in the biology since PSC characteristics depend on the dentritic tree at which APs from other neurons arrive (Deutsch and Deutsch, 1993). The dendrites and the junctions between dendrites account for weighting and spreading of PSCs as they move the information from the periphery to the neuron body. Similar scaling or weighting factors are also found and used in spiking neural network theory (Maass and Bishop, 1999), ranking the different inputs by importance. Obtaining these

weights is a daunting task and this is exactly the reason why the optimal linear filter was found first.

5.5.3.2 Optimal linear filter

Although PSC filters can be found independently from the optimal filter, the design strategy for both is initially the same. Since the interest however lies in how close to optimal a particular PSC filter can get, the optimal filter is designed and evaluated first. Details regarding optimal filter derivation and implementation can be found in Eliasmith and Anderson (2003).

The approach to finding an optimal or PSC filter is similar in that it is the best decoder for a *set* of signals. It thus necessitates that the set of input signals be defined as an *ensemble* where

$$x(t;\mathbf{A}) = \sum_{n=-(N-1)/2}^{(N-1)/2} A(\omega_n)\ e^{j\omega_n t}. \tag{5.25}$$

The set of amplitude coefficients $\mathbf{A}$ is selected from a random variable, chosen to be of a Gaussian nature in this chase. The Gaussian random variable is distributed over some bandwidth in steps of ω_n and thus represents any bandlimited signal with the given statistics.

The signal is then coded by a pair of neurons such that

$$\begin{aligned} R(t;\mathbf{A}) &= \sum_{k}^{M_{on}} \delta(t - t_k^+(\mathbf{A})) - \sum_{l}^{M_{off}} \delta(t - t_l^-(\mathbf{A})) \\ &= \sum_{i=1}^{2}\sum_{k=1}^{M} \phi_i \delta(t - t_{ik}(\mathbf{A})), \end{aligned} \tag{5.26}$$

where $\phi = 1$ for an 'on' neuron and -1 for an 'off' neuron. *Equation (5.26)* is taken as the response of the neuron pair and by minimising the mean-squared error between the input signal $x(t;\mathbf{A})$ and a decoded version of the neuron response $R(t;\mathbf{A})$ such a decoding filter $h(t)$ (see *Addendum A.3*) can be found.

In the case of the optimal filter the decoding filter is given by

$$h(\omega_n) = \frac{E\{A(\omega_n)R^*(\omega_n;\mathbf{A})\}_{\mathbf{A}}}{E\left\{|R(\omega_n;\mathbf{A})|^2\right\}_{\mathbf{A}}}, \tag{5.27}$$

where the estimate of the original signal $x(t)$ is given by

$$\hat{x}(t) = h(t) * R(t;\mathbf{A}). \tag{5.28}$$

The symbol $*$ denotes *convolution* between the optimal filter and the neuron pair response.

By defining a signal for *Equation (5.25)* an optimal filter for the decoding functional unit can be designed. From *Chapter 4* it is known that the signal to be coded by the LIF neuron is a very-low-frequency signal, since the power is calculated by squaring and then low-pass filtering the bandlimited signal with a very narrow lowpass filter (see *Section 4.5*). The bandwidth of the power signal was found to be no larger than 3 Hz (see *Section 4.6*). Firstly an optimal linear coding/decoding functional unit is designed, which works with a signal bandwidth of 5 Hz. Subsequently the change in design considerations are illustrated when the signal bandwidth is increases to 20 Hz or larger, since the smaller bandwidth implementation is a simplification of the larger one. While the optimal linear filter and accompanying PSC1 filter are not directly used in *Chapter 7* they give an insight into neural coding and estimation since a very narrow signal bandwidth greatly reduces the complexity and thus the scope of challenges a typical neural coding implementation faces.

5.5.3.3 Narrow-band signal

In *Table (5.4)* the statistics and characteristics of the narrow-band stochastic signal are given.

Table 5.4: Random variable characteristics for narrow-band stochastic signal.

RV characteristic	**Parameter**
Type	Gaussian
Bandwidth	5 Hz
Sampling frequency	5 kHz
RMS value	0.4 V

A simple Gaussian RV with a narrow bandwidth is chosen where the sampling frequency is kept large enough to allow for a better accuracy since all simulations are done in discrete-time steps.

The optimal linear filter is shown below in both the time and frequency domain.

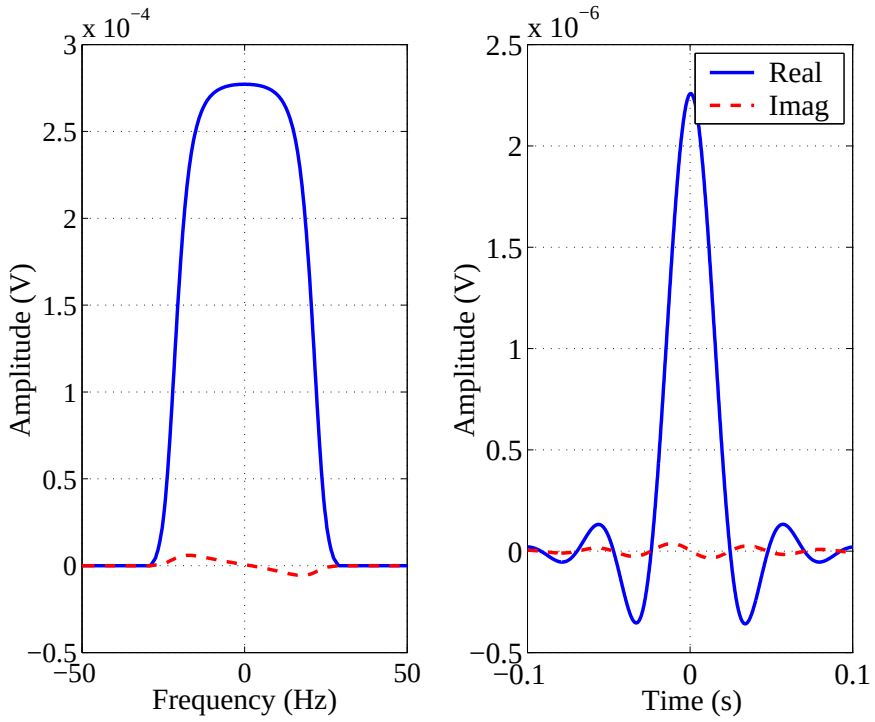

Figure 5.9: Optimal linear decoding filter for a narrow-band signal.

The optimal linear filter has both a real and an imaginary part in the frequency domain and, as expected, a narrow passband with a -3 dB cutoff of 22 Hz. In the time domain the optimal filter has the shape of a characteristic *sinc*-function and a peak amplitude of 2.3 μV. To explain why the gain of the filter and the amplitude of the impulse response are so small relative to the RMS value of the signal it has to be kept in mind that the optimal filter is defined for the *current* waveform (see *Figure 5.2*) rather than the voltage waveform where the current can lie anywhere between zero and 57.2 μA. The filter is also non-causal as it is defined for both positive and negative time.

The optimally decoded signal obtained by applying the optimal linear decoder to a random signal with statistics summarised in *Table (5.4)*, is depicted in *Figure 5.10*.

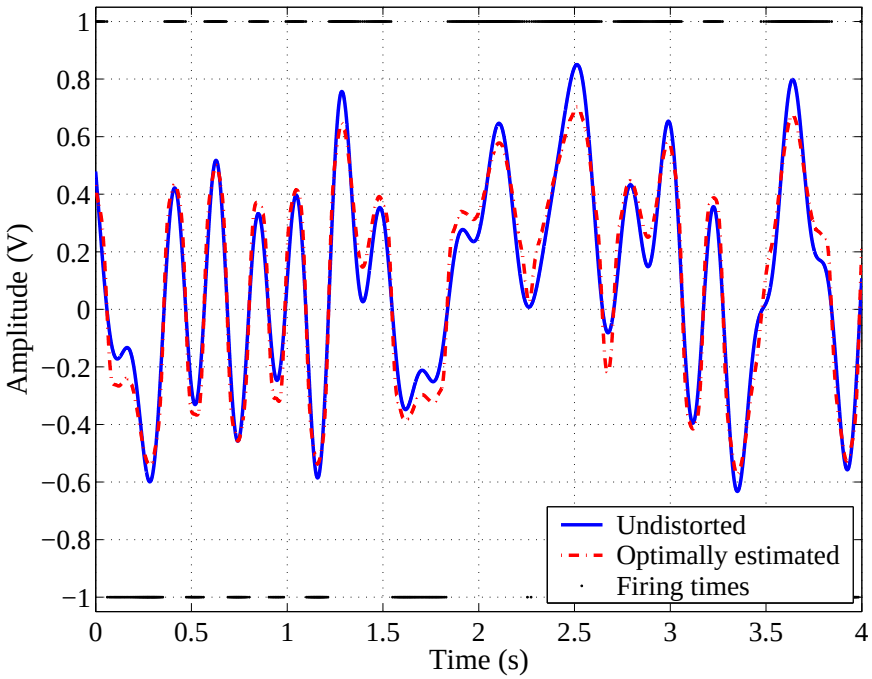

Figure 5.10: Optimally decoded signal.

In *Figure 5.10* the undistorted waveform is compared to the optimally estimated one. The dots found at an amplitude of 1 and -1 represent the occurrence of a spike at a particular point in time. The mean-squared error (MSE) defined by

$$MSE = E\left[(A_i - \hat{A}_i)^2\right], \tag{5.29}$$

where A_i is the true signal and $\hat{A}_i$ the estimated signal at sample time i is found to be 6.3 mV.

5.5.3.4 PSC filter design

Now that the optimal linear filter has been found the two PSC filter types defined in *Subsection 5.4.5* can be designed. In order to match the filtering characteristics of the PSC filter to that of the optimal linear decoder the relevant impulse, frequency and phase responses should be matched as best as possible. However, since the biology places restrictions on both the

maximum amplitude of the PSC in the time domain and the time constant, choices regarding optimisation are limited. As was mentioned in *Chapter 3* no adaptive neural decoder is implemented, limiting the design to one particular frequency response given a specific set of input signal statistics. The time constant for the LIF neuron was chosen in *Subsection 5.5.2* to be 4 ms leaving one degree of freedom, the gain of the filter. Two obvious possibilities exist when optimising for gain only, scaling the gain of the PSC filter in such a way that the DC gain matches that of the optimal filter or minimising the MSE of the decoded signal to that of the true signal by choosing the best filter gain.

The result of matching the impulse response areas of the optimal linear and PSC filters is depicted in *Figure 5.11*.

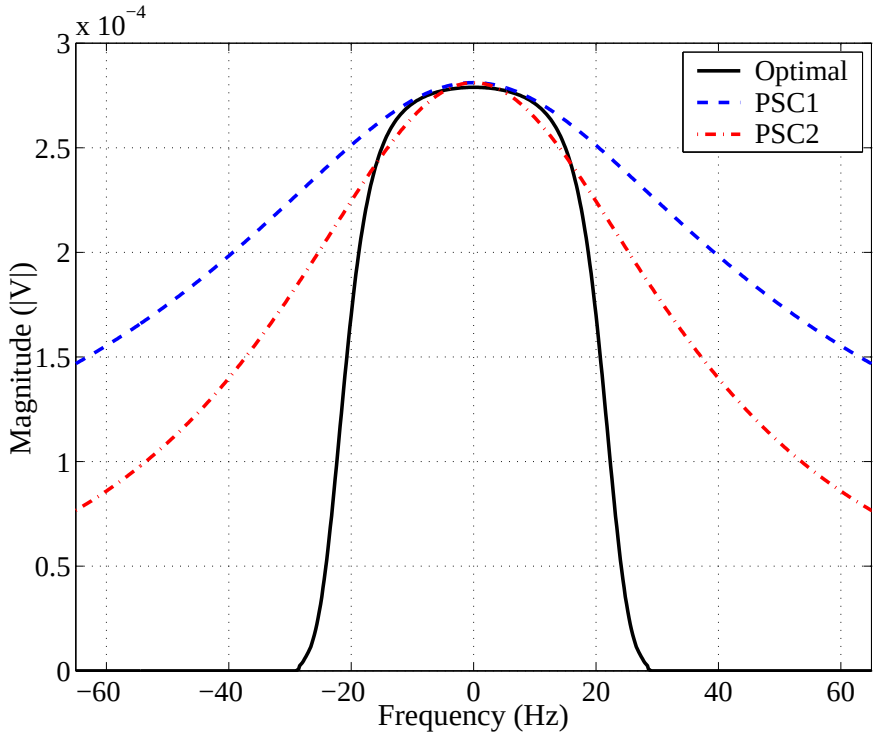

Figure 5.11: Optimised PSC filters by matching DC gain to that of optimal filter.

The PSC1 filter matches the optimal filter well in the -15 to 15 Hz band, compared to the PSC2 filter. The stopband attenuation of the PSC2 filter is however much closer to that of the optimal linear filter compared to the PSC1 filter, which can be attributed to the steeper roll-off characteristics of PSC2 filter.

The second method of minimising the MSE between the true and estimated signal is easily

implemented since only the overall gain of the PSC filter can be scaled. This can be attributed to the fact that different gains for the different frequencies stay constant in respect with one another, allowing only an optimal gain offset, which translates the frequency response along the gain axis.

The result of such scaling is shown in *Figure 5.12*.

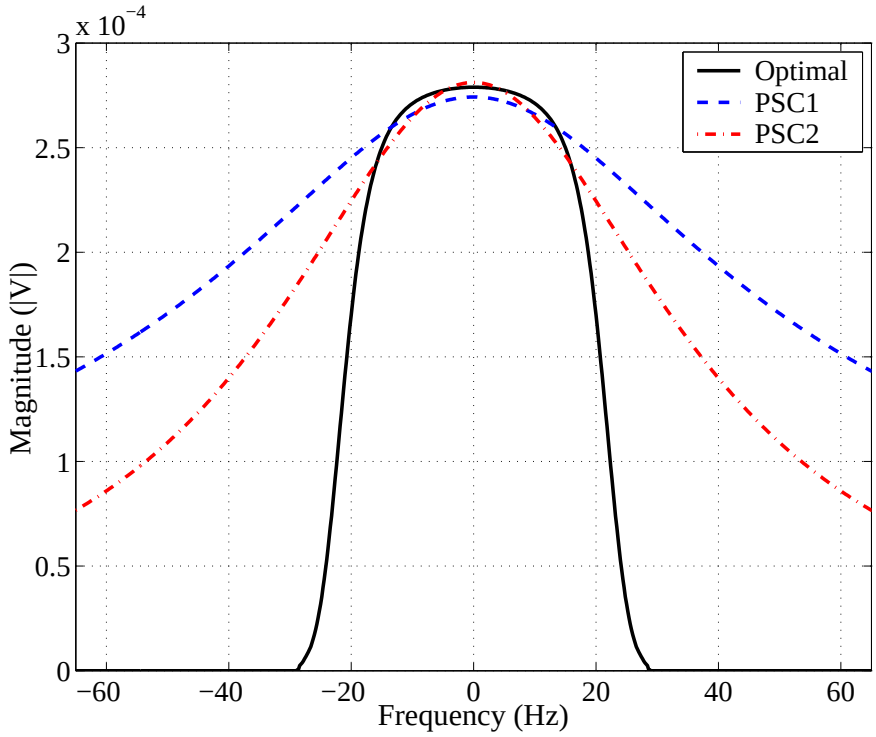

Figure 5.12: Optimised PSC filters by minimising the MSE.

Compared to *Figure 5.11* only very little difference (0.8e-3 dB) exists between the two PSC2 filters as far as the frequency response is concerned. There however exists a 0.1 dB difference between the gain of the PSC1 filter found in *Figure 5.11* and that of the filter shown in *Figure 5.12*. Thus in the case of the second PSC filter from a MSE point of view, scaling the filter in such a way as to match the DC gain to that of the optimal linear filter was a good one. It is however evident that since a change in the optimal linear filter bandwidth also compromises the validity of such a choice for the second PSC filter, the best way would be to scale the two PSC filter in such a way as to minimise the MSE.

5.5.3.5 Wide-band signal

In the previous sections the optimal linear filter and accompanying optimised PSC filters were designed for a narrow-band signal relevant for this problem. However, what happens if the signal has a larger bandwidth? In this section this question was briefly addressed and the implications thereof discussed. In *Table (5.5)* the statistics and characteristics of the narrow-band stochastic signal are given.

Table 5.5: Random variable characteristics for wide-band stochastic signal.

RV characteristic	**Parameter**
Type	Gaussian
Bandwidth	20 Hz
Sampling frequency	5 kHz
RMS value	0.4 V

A simple Gaussian RV with a 20 Hz bandwidth is chosen.

The optimal linear filter is shown below in both the time and frequency domain.

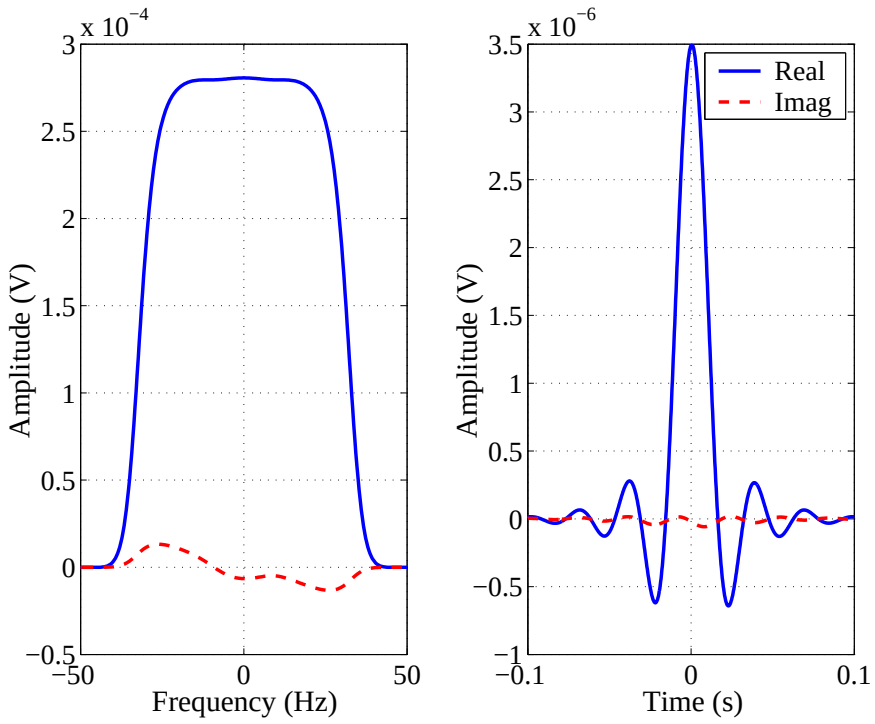

Figure 5.13: Optimal linear decoding filter for a wide-band signal.

The passband of the optimal linear filter is, as expected, wider than the one in *Figure 5.9* with a -3 dB cutoff of 32 Hz (11 Hz wider). In the time domain, the optimal filter also has the shape of a characteristic *sinc*-function but a peak amplitude of 3.5 μV. The impulse response width is however, as expected, narrower than the one in *Figure 5.9*.

The wider frequency response now poses a problem when trying to fit a PSC filter to it. As before it is attempted to fit the PSC frequency responses to the optimal linear filter, by firstly matching the DC gain of the PSC filters to that of the optimal linear filter.

The results are shown in *Figure 5.14*.

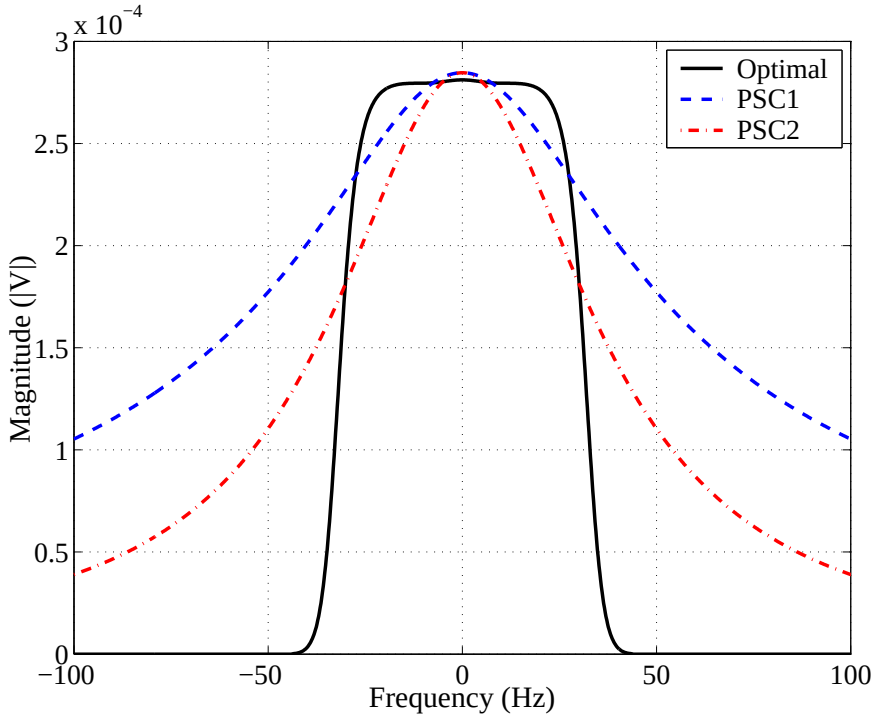

Figure 5.14: Optimised PSC filters by matching DC gain to that of optimal filter for wide-band signal.

Even the wider PSC1 filter is now simply not wide enough to match the passband of the optimal filter. Since the time constant is chosen to be fixed, the passband of the PSC filter remains fixed.

If the alternative method of scaling the gain of the PSC filter by minimising the MSE between true and estimated signal is chosen, a similar gain is found for both PSC1 and PSC2 as *Figure 5.15* illustrates.

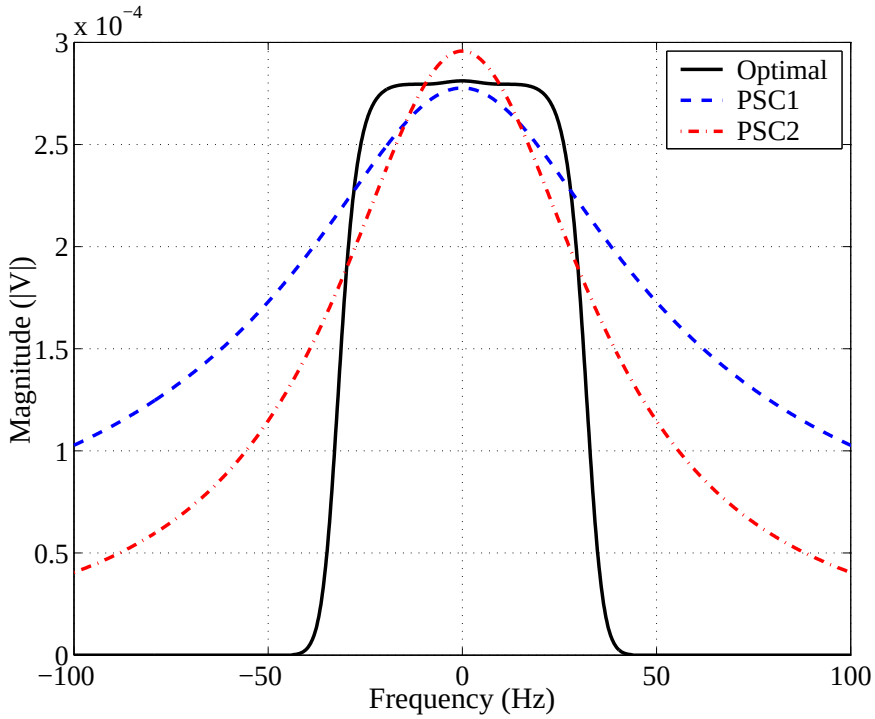

Figure 5.15: Optimised PSC filters by minimising MSE for wide-band signal.

It is evident that as the bandwidth of the signal to be coded increases, the "best fit" for the PSC filters remains the same, but that the MSE will increase. It can be deduced from this that utilising PSCs for decoding purposes limits the bandwidth of the signal to be coded and decoded. As will be seen in the next few sections, the PSC2 filter competes very favourable with the optimal filter for narrow-band input signals when a biologically plausible time constant is chosen.

5.5.3.6 PSC filter application

The two optimisation methods for the narrow-band input signal case were used to optimise the weight of the PSC filters in the MSE sense and the decoding PSC filters applied to a decoding problem.

Table (5.6) summarises the performances of the optimal linear decoder and the PSC filters in the MSE sense.

Table 5.6: MSE performance of various filters.

Performance parameter	**Optimal filter**	**PSC1 filter**	**PSC2 filter**
MSE (V^2)	0.0063	0.0187	0.0070

While the optimal filter fared the best over a time interval of 4 s, the second PSC filter performed significantly better than the PSC1 filter. In *Figure 5.16* the signal estimated by the first PSC filter is shown.

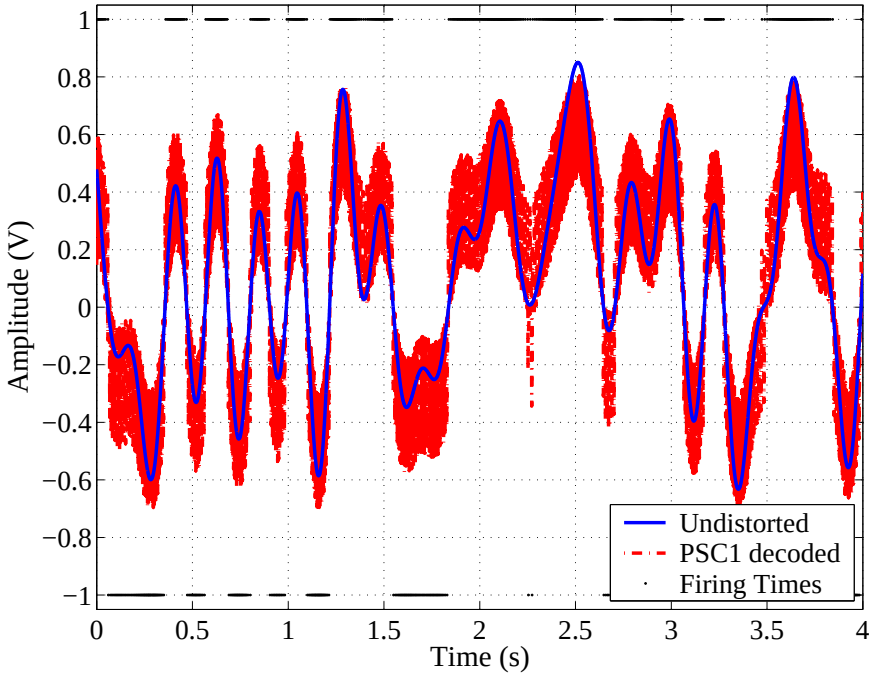

Figure 5.16: Decoding performance of PSC1 filter.

In *Figure 5.17* the signal estimated by the PSC2 filter is shown along with the true signal.

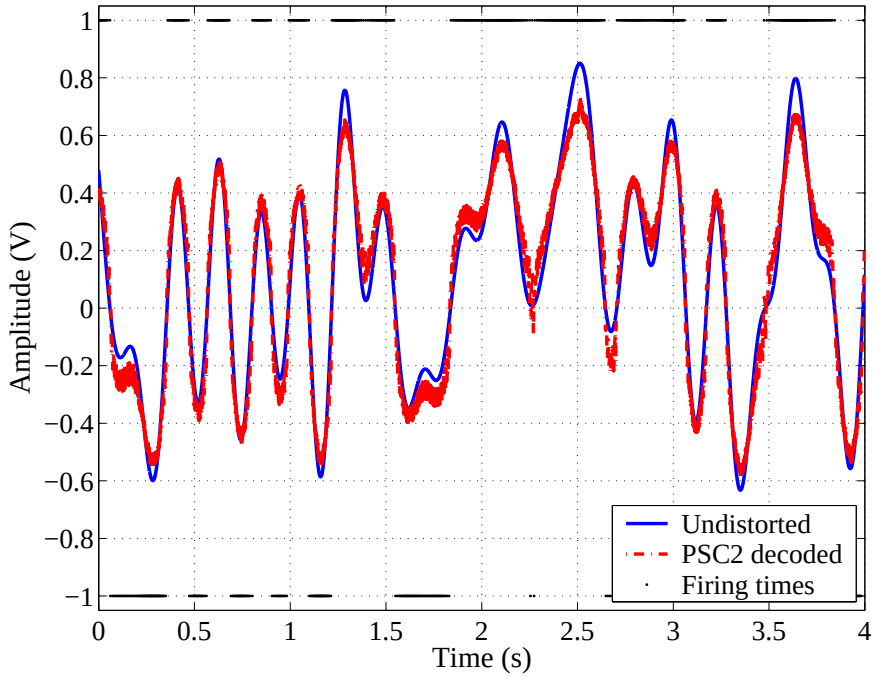

Figure 5.17: Decoding performance of PSC2 filter.

The estimated signal in *Figure 5.17* looks significantly better than the one shown in *Figure 5.16*.

5.5.3.7 Decoding in the presence of noise

To evaluate how the decoding filters perform in the presence of noise the noise model specified in *Subsection 5.3.2* is now implemented. While the PSC filters were scaled optimally in the MSE sense in the *absence* of noise no additional scaling is introduced to optimise for a particular type of noise. The two noise sources spike-time jitter and the dropping of spikes was applied in a similar way to that of Bialek et al. (1991). The MSE for the noiseless case is compared to where either spike-time jitter was introduced or spikes were dropped.

The results are summarised in *Table (5.7)*.

Table 5.7: PSC filter performance in the presence of noise.

Noise	**MSE (PSC1 filter)** (V^2)	**MSE (PSC2 filter)** (V^2)
none	0.0187	0.0070
$\sigma = 0.8$ ms	0.0206 (-0.43 dB)	0.0074 (-0.25 dB)
10 % dropped spike	0.0242 (-1.12 dB)	0.0109 (-1.91 dB)

From *Table (5.6)* it can be seen that the PSC2 filter is 4.26 dB better in decoding the random narrow-band signal than the PSC1 filter when no noise is present and with a similar difference (4.44 dB) in the case when jitter noise, with a standard deviation of 0.8 ms is added. In the case of randomly dropping 10 % of the spikes present, the PSC2 filter performs only 3.47 dB better in MSE sense than the PSC1 filter.

In *Figure 5.18* and *Figure 5.19* the estimated input signal is compared to the true signal in the presence of spike-jitter noise over an arbitrary 4 s interval at a sampling frequency of 4 kHz for the two different PSC filters.

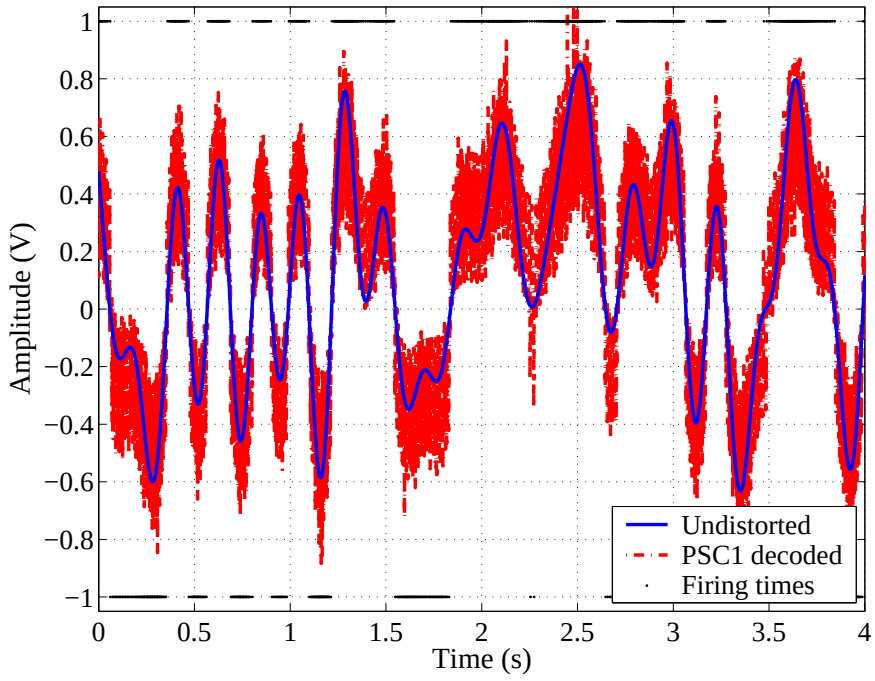

Figure 5.18: PSC1 filter decoding performance in the presence of spike-jitter noise with a standard deviation of 0.8 ms.

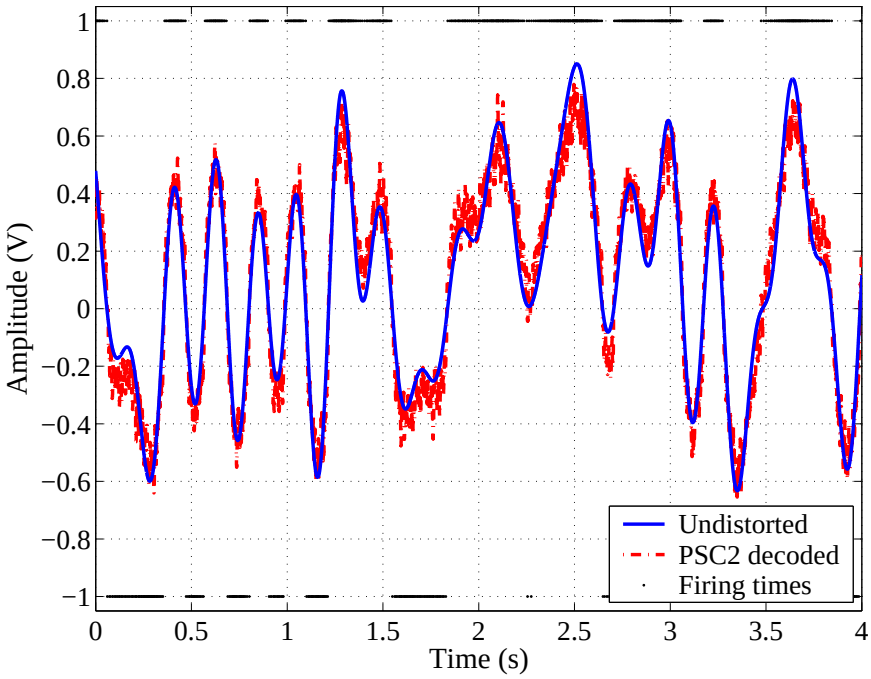

Figure 5.19: PSC2 filter decoding performance in the presence of spike-jitter noise with a standard deviation of 0.8 ms.

In *Figure 5.20* and *Figure 5.21* the estimated input signal is compared to the true signal when 10 % of its spikes are randomly lost over the unreliable channel due to noise.

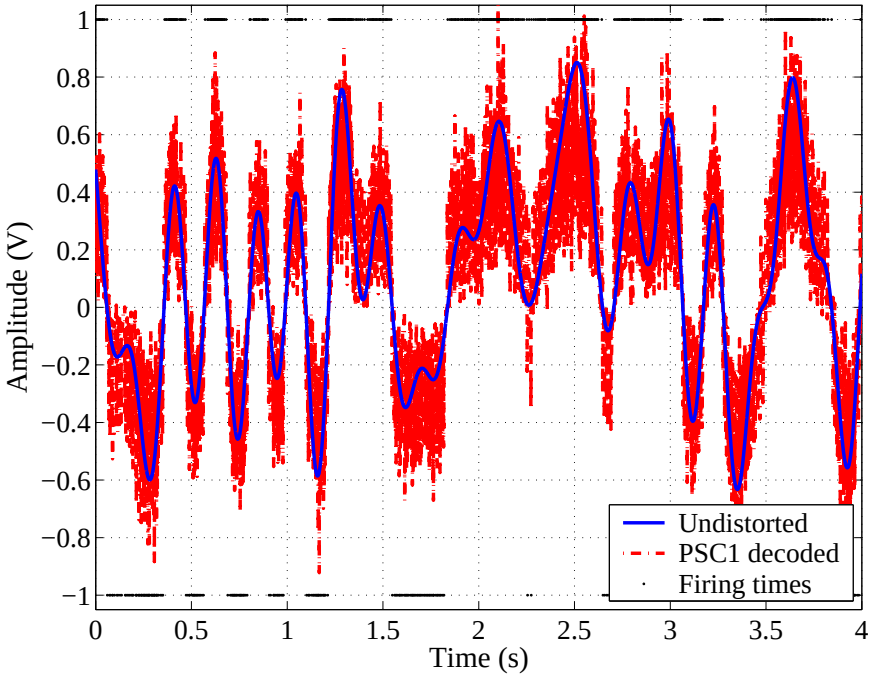

Figure 5.20: PSC1 filter decoding performance in the presence of 10 % spike loss.

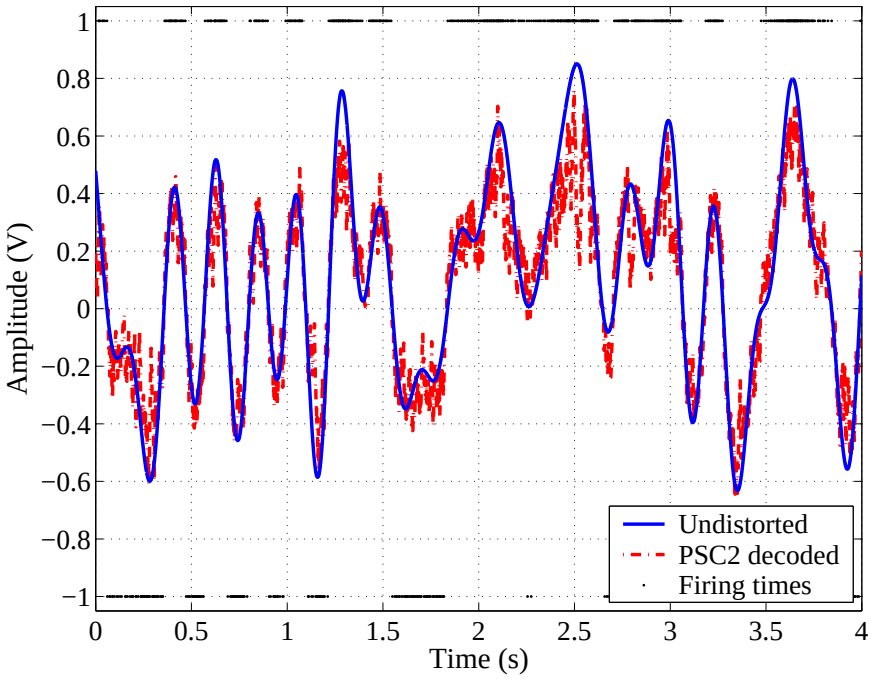

Figure 5.21: PSC2 filter decoding performance in the presence of 10 % spike loss.

The MSE of the two PSC decoders was evaluated as a function of the standard deviation in spike-time jitter and the results are depicted in *Figure 5.22*.

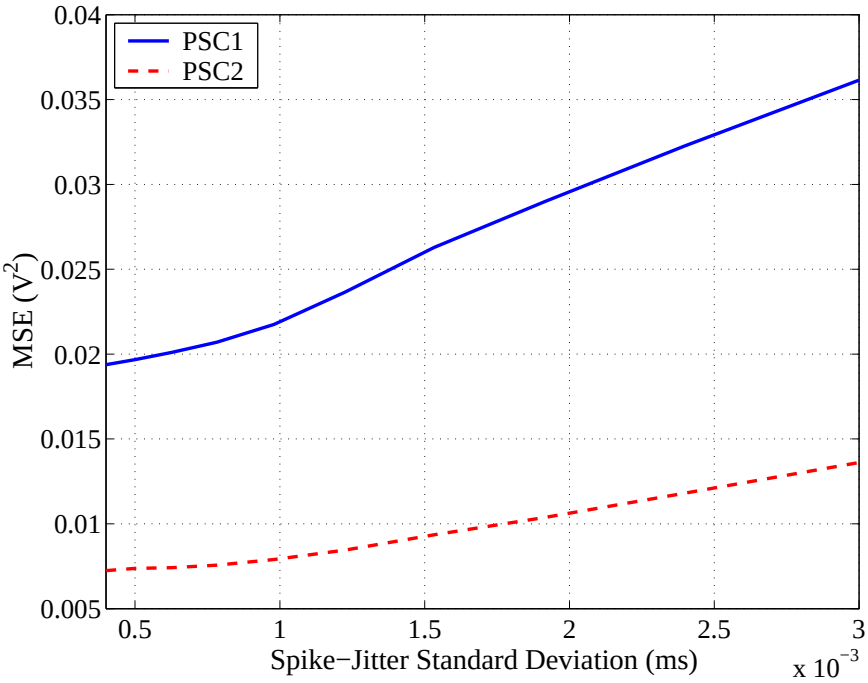

Figure 5.22: MSE as a function of spike-time jitter for PSC1 and PSC2 filters.

The MSE at a particular spike-jitter variance was obtained by statistically averaging over 5 independent intervals of 4 s each, sampled at 4 kHz ($80 \cdot 10^3$ samples for statistical approximation). As the standard deviation of the spike-jitter noise increases from 0.4 ms to 3 ms, the MSE increases for both PSC1 and PSC2 decoding filters. Above a standard deviation of 1 ms the MSE due to PSC1 decoding increases almost twice as fast as that of the PSC2 filter (MSE difference at 0.4 ms is 0.0121 and at 3 ms 0.0225.)

Similarly to *Figure 5.22* the effect that changing the number of dropped spikes has on the MSE of the decoder is shown in *Figure 5.23*.

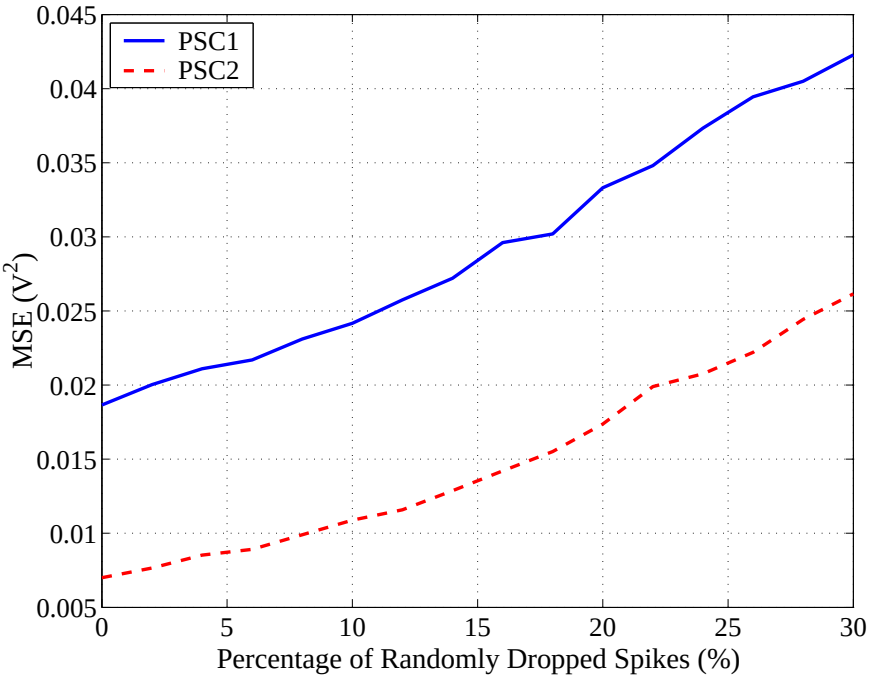

Figure 5.23: MSE as a function of dropped spikes for PSC1 and PSC2 filters.

The same statistical averaging parameters as for the spike variance case was used to obtain *Figure 5.23*. The MSE increases linearly as a function of dropped spikes and the error doubles after 25 % of the spikes are lost due to noise. In this case averaging over eighty thousand samples gives a good approximation of the trend the MSE curve follows when an increased number of spikes are dropped. If the MSE is computed over a larger set of random samples the curve is expected to be smoother. Due to a time consuming spike generation process via the convolution integral of *Equation (5.12)* the number of random samples averaged over statistically was kept to a minimum.

5.6 DISCUSSION AND CONCLUSION

In this chapter a coder/decoder based on a biologically plausible estimator, the PSC filter, was successfully designed, implemented and evaluated. In *Section 3.8* no particular code was chosen. Instead, a biologically plausible neuron model, namely the LIF neuron, was chosen. As was mentioned in *Section 3.3* a rate code conveys only the average rate by determining the average number of spikes in a particular window size. From the phase response of the PSC filters given in *Figure 5.6* and *Figure 5.8* it can be deduced that a non-linear phase is introduced to the estimated signal, which was removed for the sake of convenience in the above evaluation. What is interesting to note is that as the time constant of the PSC filters increase the bandwidth of the filter decreases, which in fact is nothing more than increasing the averaging window just mentioned. As the time constant is increased the averaging window used by a rate code increases. This is where the crux of picking a neural code lies. While amplitude, frequency and phase information are present in the coded signal, picking a particular decoding technique limits the observable degree of freedom of the signal. The choice of a biologically plausible time constant of 4 ms still acts as an averaging window, yet largely retains amplitude, frequency and phase (delayed) information. It can be concluded that according to the definition of rate and temporal codes a *combination* of both codes is used. By changing coding and decoding parameters it was chosen to ignore certain parts of the information or to include all of it, therefore moving between a rate code and a time code.

Another issue that was addressed was the scaling of the PSC filters. Since the normalised PSC filters of *Equation (5.9)* and *Equation (5.10)* are not scaled to an amplitude suitable for proper decoding, a method was attained to scale the PSCs in order to minimise the MSE between the original input signal and the estimated one. This was done by first finding an optimal filter according to Eliasmith and Anderson (2003) from which the PSC filter for the two different kernels, PSC1 and PSC2, was found. Since it was chosen to implement a fixed time constant the only other PSC filter variable that could be changed was the amplitude (gain). It was discovered that minimising the MSE rendered a scaled PSC filter capable of decoding an ensemble of input signals as best as possible (in the MSE sense) given a fixed time constant. It was also found that for a narrow-band signal, as in this case (see *Section 4.6*), the PSC2 filter performs well compared to the optimal linear filter since the frequency response for such a narrow-band signal compares quite favourable to that of the PSC2 filter with steep roll-off characteristics. Matching the surface areas of the PSC2 filter to that of the optimal filter is similar to matching the DC gains of the two filters. An analogous result is

obtained when the optimum weighted PSC2 filter is found in the MSE sense.

The PSC1 filter on the other hand with its shallower roll-off characteristics fits the frequency response profile of the optimal linear decoder less favourable and even with optimal weighting of the filter in the MSE sense, the PSC1 filter performs 4.26 dB worse than the PSC2 filter for the same input signal. In general, the PSC2 filter performs in the order of 3.47 - 4.44 dB better than the PSC1 filter.

Both PSC1 and PSC2 decoders are sensitive to spike-jitter noise and the loss of spikes due to unreliable retransmissions between successive neurons. However, both filters are less vulnerable to jitter noise than the dropping of spikes since the loss of spikes decreases the overall power contained within the spike train as compared to spike-jitter noise where the overall power is maintained but randomly redistributed. Since the MSE, defined in *Equation (5.29)*, is directly influenced by a change in amplitude and hence in power the results obtained can be intuitively substantiated.

Since the PSC2 filter is a biologically more plausible filter than the optimal linear filter and the PSC1 filter, it makes for a good choice for a decoding filter. Not only is the filter causal and physically realisable but for a time constant of 4 ms and a narrow-band signal specified in *Table (5.4)* performs unexpectedly well compared to the optimal linear filter (0.46 dB worse). Since the PSC2 filter is easier to implement and computational less intensive it is the best choice that can be made in choosing a matching decoding functional unit for implementation purposes in *Chapter 7*.

CHAPTER 6

ESTIMATOR MODELS

6.1 AIM OF THIS CHAPTER

The aim of this chapter is to find an estimator based on the information gained on the auditory system model specified and designed in the preceding chapters. An estimator provides a tool to extract information from a noisy environment, given the approximate knowledge of the statistics of the underlying process. Since the entire process has been designed in the preceding chapters *Chapter 4* to *Chapter 5* and is thus known, an estimator (optimal or suboptimal) can be found to estimate the state variables of the auditory system.

In particular, an implementation of the well-established *Kalman filter* (Van Trees, 1968-1971; Anderson and Moore, 1979; Bozic, 1979; Gelb, 1984; Sorenson, 1985; Lewis, 1986; Chui and Chen, 1987; Brown and Hwang, 1997; Grewal and Andrews, 2001) is pursued in order to enable estimation of the decoded signal $z(t)$ in *Figure 6.1*a.

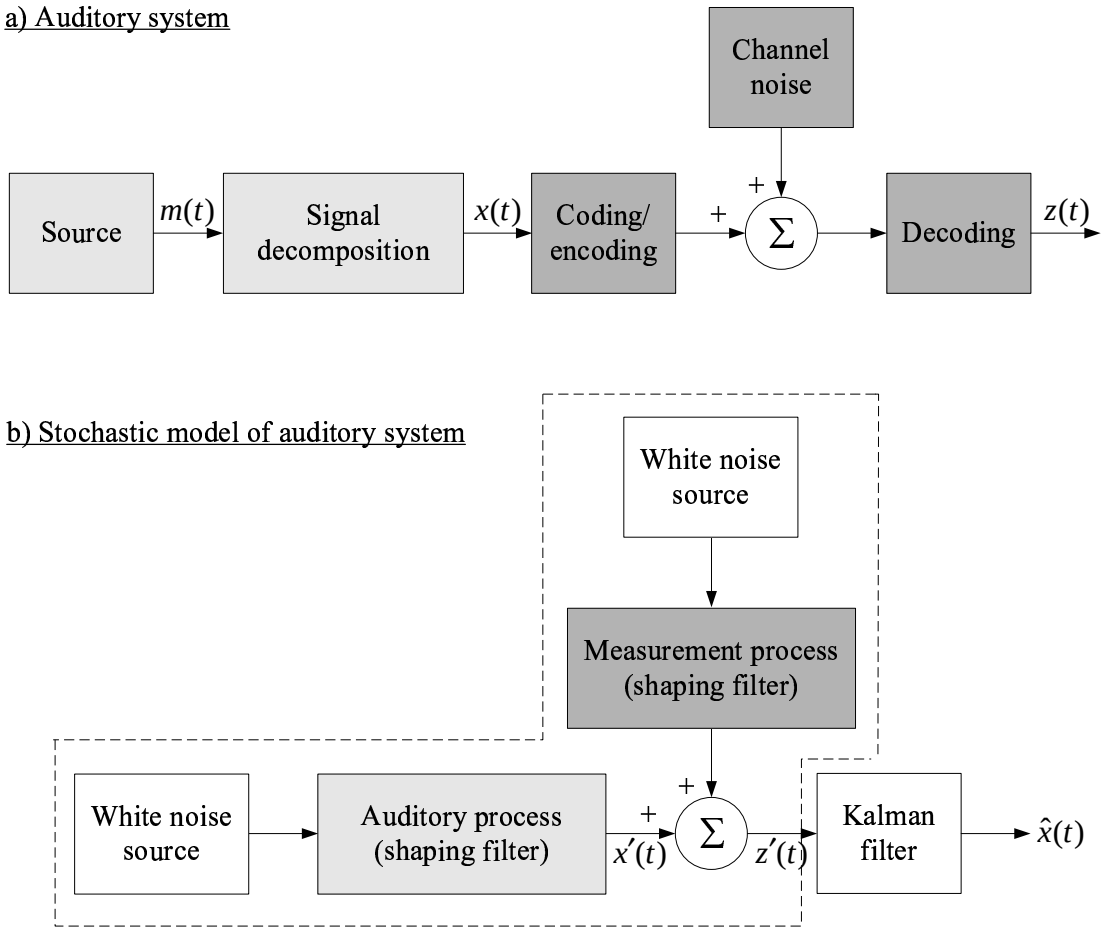

Figure 6.1: Auditory system model a) designed in preceding chapters and stochastic model for a) shown in b) will be designed in this chapter. $x'(t)$ and $z'(t)$ are the stochastic equivalents of $x(t)$ and $z(t)$. (Reproduced from *Figure 1.9*)

This chapter provides all information to design, implement, initialise and evaluate the particular filter.

6.2 ESTIMATION PROBLEM

6.2.1 Power Profile Estimation Problem

The aim of the estimator is to estimate the amount of power found in a particular frequency band, as specified in *Section 4.5*, given a noisy measurement of the estimated signal. The estimated signal is found after decoding of spikes has taken place by a PSC filter (designed and evaluated in *Chapter 5*). The idea is to optimally (in the MMSE sense) reconstruct the power profile across the entire bandwidth of the system. Since the power profile contains most information of the original signal that was received at the filter bank, it can be used by higher hierarchial level systems for further processing as was discussed in *Figure 1.10*. In this case, finding a robust estimate of such a power profile is the main priority. Since the information within the power profile is severely corrupted by noise when transmitted across the neural channel (see *Section 5.3*), it demands a robust estimator.

One of the most well-established recursive estimators is the Kalman filter. Given the approximate underlying process and a noisy measurement it can reliable estimate the underlying states of any system, provided that it is observable, controllable and stable as will be discussed later.

6.3 KALMAN FILTERING

In *Section 1.4* the approach implementing Kalman filtering was briefly discussed and outlined. It is assumed at this point that the reader has little knowledge of the Kalman filter design methodology, warranting a systematic and detailed design approach with the relevant background theory. The reader more familiar with Kalman filters is however still urged to read the following sections since the design information is contained within the sections. The function of the Kalman filter is discussed next.

In order to estimate the state of a dynamic linear stochastic system the Kalman filter implements an optimal, in the minimum mean-squared error (MMSE) sense, recursive and linear filter with time-varying gain. If the system that generates the measurement signal is linear and the noise added to the signal is Gaussian, then the Kalman filter is the optimal MMSE estimator among both linear *and* non-linear filters (Chui and Chen, 1987; Brown

and Hwang, 1997; Grewal and Andrews, 2001; Lewis, 1986; Hammarberg, 2002). The system, in state-space formulation, describes how the measurements are related to a set of inputs that may be purely stochastic, deterministic or a combination of the two. One of the greatest advantages of the Kalman filter is its recursive implementation, which requires only the knowledge of the previous state to make an estimate thus greatly reducing memory requirements and increasing the computational efficiency. Both continuous and discrete-time implementations of the Kalman filter algorithm exist, which allow for inherently continuous, sampled-continuous or inherently discrete processes. A suboptimal version of the Kalman filter, the extended Kalman filter (EKF) applied to inherently non-linear systems, is used in this dissertation.

As was noted previously, the first step of designing a Kalman filter is to find a state-space model of the underlying process. The state-space equations are defined next.

6.4 STATE-SPACE EQUATIONS

The basic Kalman filter estimator parameters can only be estimated optimally once a state-space model has been derived for the underlying process.

Any nth-order differential equation describing some system has to be broken down into n first-order differential equations, which form the set of state (system) equations and one measurement equation given below.

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) \\ \mathbf{y}(t) &= \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t). \end{aligned} \tag{6.1}$$

In *Equation (6.1)*, $\mathbf{u}(t)$ is assumed to be a deterministic input driving a static system, since the system attributes matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{D}$ are independent of time. In this dissertation a time-varying matrix in the continuous-time is represented as $\mathbf{A}(t)$ and in the discrete-time by a subscript such that $\mathbf{A}_k$. Thus the system attribute matrices in *Equation (6.1)* would indicate a *static* system. Since the nomenclature for the system attribute matrices vary, a formal definition is required.

The most widely used names are summarised in *Table (6.1)*.

Table 6.1: State-space equation variables.

Symbol	Variable name
A	System matrix
B	Input coupling matrix
C	Measurement sensitivity matrix
D	Output coupling matrix
u	Input or control vector
x	State vector
$\dot{\mathbf{x}}$	Derivative of state vector
y	Output vector

Symbols of variables that are defined at a later stage and are mainly used in conjunction with the Kalman filter are given by *Table (6.2)*.

Table 6.2: Other state-space equation variables.

Symbol	Variable name
$\boldsymbol{\Phi}$	Discrete-time state transition matrix
$\boldsymbol{\Gamma}$	Discrete-time process noise coupling matrix
H	Discrete-time measurement sensitivity matrix
$\boldsymbol{\omega}$	White Gaussian process noise vector
w	Coloured process noise vector
v	Measurement noise vector

Since some processes are inherently discrete in time and information regarding the system variables are only available at certain discrete instances in time, a discrete-time representation of *Equation (6.1)* could be required. The n first-order differential equations then become n first-order difference equations of the form

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{A}_k\mathbf{x}_k + \mathbf{B}_k\mathbf{u}_k \\ \mathbf{y}_k &= \mathbf{C}_k\mathbf{x}_k + \mathbf{D}_k\mathbf{u}_k. \end{aligned} \tag{6.2}$$

A set of differential or difference equations, for a particular system for which the state-space equations are required, can be obtained in more than one way and generally depend on the type of information that is available to the Kalman filter designer. The most straightforward approach is to find the linear transfer function of the system. The state-space equations have to be derived for the stochastic equivalent model of the auditory system model as is discussed below.

6.5 STATE-SPACE MODELLING OF AUDITORY SYSTEM

The stochastic model for which the state-space equations have to be derived is shown in *Figure 6.2*. This will be the internal model of the Kalman filter.

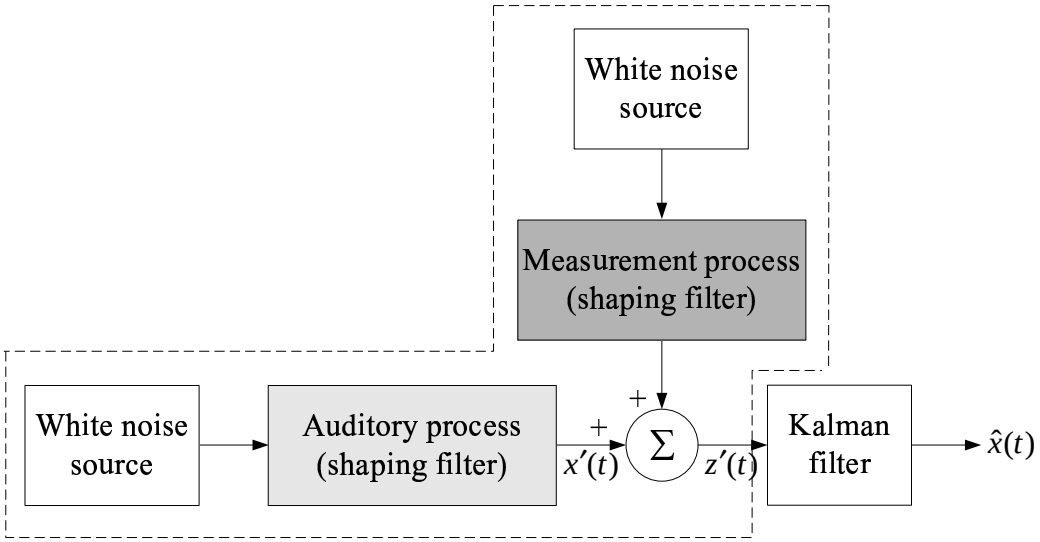

Figure 6.2: Stochastic equivalent model for auditory system model.

The auditory system model, containing the source and signal decomposition block (see *Figure 6.1*a) and driven by a white-noise source, is derived first and written in state-space form of *Equation (6.2)*. Subsequently the measurement process model, containing the coding, channel, noise and decoding block, is derived. However, a formal definition for the Kalman filter is appropriate.

6.6 DISCRETE-TIME KALMAN FILTER

Combining the Kalman filter, first introduced by the famous paper by Kalman (1960), with the advances in digital computer technology makes it possible to easily implement its recursive solution. The Kalman recursive equations consist of system and measurement equations and will be described in more detail here.

6.6.1 Discrete-Time Kalman Filter State-Space Equations

A random process to be estimated can be modelled by a n-input, state-equation of the form

$$\mathbf{x}_{k+1} = \mathbf{\Phi}_k \mathbf{x}_k + \mathbf{\Gamma}_k \boldsymbol{\omega}_k + \mathbf{B}_k \mathbf{u}_k, \tag{6.3}$$

where $\mathbf{\Phi}_k$ (n x n) is the state transition matrix, $\mathbf{x}_k$ (n x 1) the vector containing all state variables and $\boldsymbol{\omega}_k$ (n x 1) the zero mean, white Gaussian process noise vector coupled by the process noise coupling matrix $\mathbf{\Gamma}_k$ (n x n). $\mathbf{u}_k$ (n x 1) is a deterministic input, which is set to zero since no deterministic input is required in this case.

The measurement of the process is assumed to occur in discrete-time steps given by the m-output linear relationship

$$\mathbf{z}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k, \tag{6.4}$$

where $\mathbf{H}_k$ (m x n) is the measurement sensitivity matrix and $\mathbf{v}_k$ (m x 1) the measurement noise. A block diagram representation of *Equation (6.3)* and *Equation (6.4)* is given in *Figure 6.3*a.

The noise process driving the system and measurement equations is defined to adhere to the following *covariance* matrix conditions

$$E\left[\boldsymbol{\omega}_k\boldsymbol{\omega}_i^T\right] = \begin{cases} \mathbf{Q}_k, & \text{i=k} \\ 0, & \text{i}\neq\text{k}, \end{cases} \tag{6.5}$$

$$E\left[\mathbf{v}_k\mathbf{v}_i^T\right] = \begin{cases} \mathbf{R}_k, & \text{i=k} \\ 0, & \text{i}\neq\text{k}, \end{cases} \tag{6.6}$$

$$E\left[\boldsymbol{\omega}_k\mathbf{v}_i^T\right] = 0, \quad \text{for all k and i,} \tag{6.7}$$

where $\mathbf{Q}_k$ and $\mathbf{R}_k$ are defined as the covariance matrices associated with $\boldsymbol{\omega}_k$ and $\mathbf{v}_k$ respectively, at a particular sampled interval k in time.

6.6.2 Discrete-Time Kalman Filter Equations

The discrete-time Kalman filter, being a minimum covariance error estimator, can be described by the following equations.

The Kalman gain matrix is:

$$\mathbf{K}_k = \mathbf{P}_k(-)\mathbf{H}_k^T\left[\mathbf{H}_k\mathbf{P}_k(-)\mathbf{H}_k^T + \mathbf{R}_k\right]^{-1}. \tag{6.8}$$

The state estimation observable update is:

$$\hat{\mathbf{x}}_k(+) = \hat{\mathbf{x}}_k(-) + \mathbf{K}_k\left[\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_k(-) - \mathbf{H}_k\mathbf{B}_{k-1}\mathbf{u}_{k-1}\right]. \tag{6.9}$$

The error covariance update is:

$$\mathbf{P}_k(+) = \left[\mathbf{I} - \bar{\mathbf{K}}_k\mathbf{H}_k\right]\mathbf{P}_k(-). \tag{6.10}$$

The state estimation extrapolation is:

$$\hat{\mathbf{x}}_{k+1}(-) = \mathbf{\Phi}_k \hat{\mathbf{x}}_k(+) + \mathbf{B}_k \mathbf{u}_k. \tag{6.11}$$

The error covariance extrapolation is:

$$\mathbf{P}_{k+1}(-) = \mathbf{\Phi}_k \mathbf{P}_k(+) \mathbf{\Phi}_k^T + \mathbf{\Gamma}_k \mathbf{Q}_k \mathbf{\Gamma}_k^T, \tag{6.12}$$

where all symbols are defined below.

The Kalman filtering process can be described as follows: The Kalman filter is initialised by specifying an initial state estimate $\hat{\mathbf{x}}_0(-)$, where the $(-)$ indicates the *a priori* and $(+)$ the *a posteriori* estimate. It is assumed that the statistics of the state estimate at time t_0 are known and are contained in the error covariance matrix $\mathbf{P}_0(-)$. Based on the statistics of the measurement noise given by $\mathbf{R}_0$, the Kalman filter gain matrix $\mathbf{K}_0$ can be found by solving for the Kalman gain matrix (*Equation (6.8)*). The Kalman filter gain minimises the mean-squared estimation error and can now be used to find the *a posteriori* estimate $\hat{\mathbf{x}}_0(+)$, using *Equation (6.9)*, based on the error between the measurement $\mathbf{z}_0$, the estimated measurement $\mathbf{H}_0\hat{\mathbf{x}}_0(-)$ and, if present, the deterministic input $\mathbf{B}_0\mathbf{u}_0$. The *a posteriori* error covariance matrix $\mathbf{P}_0(+)$ is found by reflecting the changes applied by the Kalman gain to the *a priori* error covariance matrix $\mathbf{P}_0(-)$ (see *Equation (6.10)*). The Kalman filter can now project ahead and estimate the state variables (*Equation (6.11)*) and error covariance matrix (*Equation (6.12)*) at t_1, based on the *a posteriori* estimate and statistics of the process noise contained in $\mathbf{Q}_0$. The process is subsequently repeated.

Equation (6.11) and *Equation (6.9)* are visually shown in the block diagram of *Figure 6.3*.

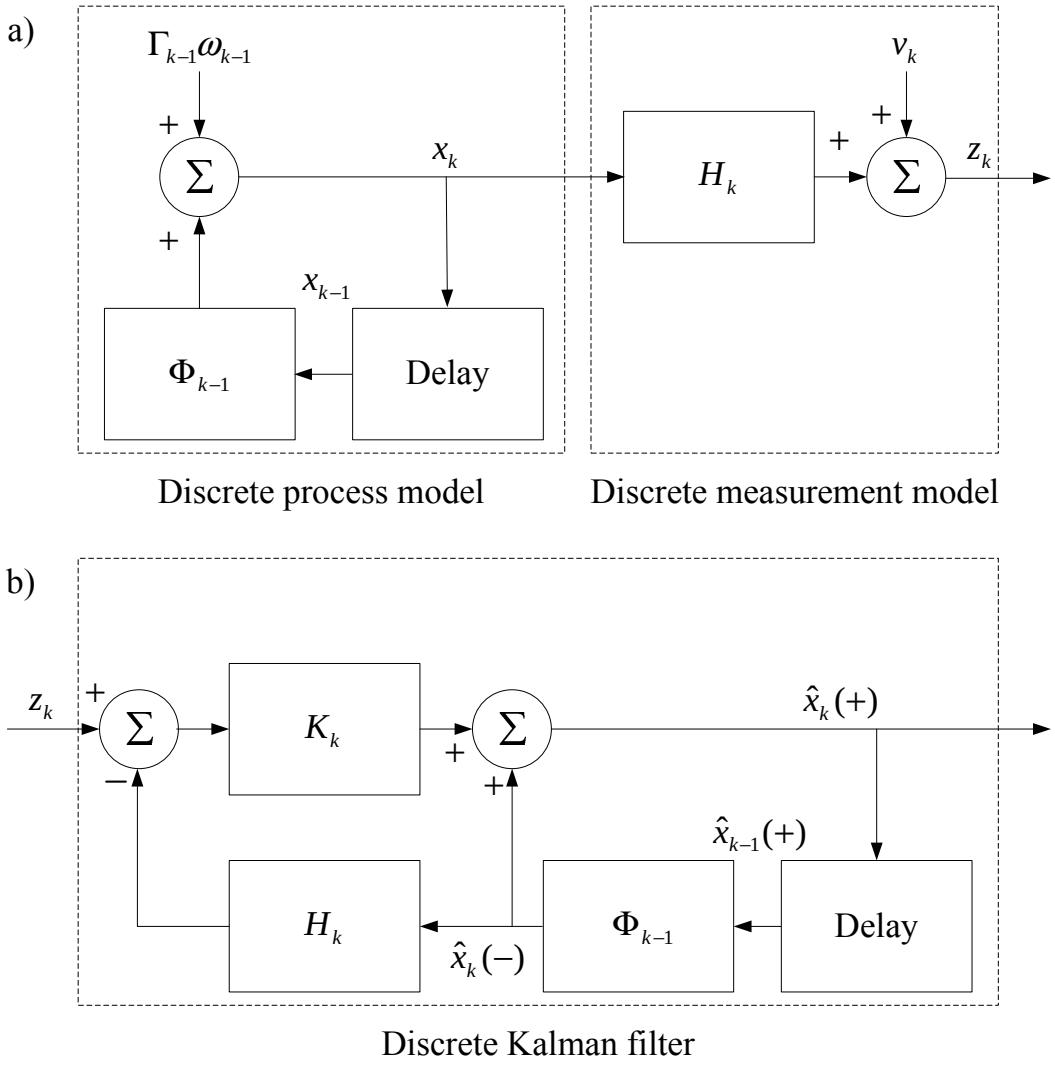

Figure 6.3: Block diagram of a) system, measurement model and b) discrete-time Kalman filter.

From the Kalman filter equations (*Equation (6.11)* to *Equation (6.8)*) and the block diagram above it should now be clear to the reader why an *accurate* description of the auditory model is required. The Kalman filter contains an internal model of the process that generates the measurements through matrices $\mathbf{\Phi}_k$ and $\mathbf{H}_k$. Since the Kalman filter relies on a good model of the underlying process found in the matrices $\mathbf{\Phi}_k$ and $\mathbf{H}_k$ and on correct initialisation of the state variables $\mathbf{x}_k$ and error covariance matrix $\mathbf{P}_k$, any errors in the matrices $\mathbf{\Phi}_k$ and $\mathbf{H}_k$ not compensated for by process noise $\boldsymbol{\omega}_k$ or measurement noise $\mathbf{v}_k$ will lead to divergence (Gelb, 1984; Brown and Hwang, 1997; Grewal and Andrews, 2001) of the Kalman filter. Correct and consistent initialisation is addressed in the next section.

6.6.3 Correct and Consistent Initialisation

Before using the Kalman filtering equations *Equation (6.11)* to *Equation (6.8)*, the initial state estimate and its covariance matrix have to be set up in order to achieve optimal (in the MMSE sense) performance. If initial values are uncertain, measurements made can be used to find the MMSE state estimates at the starting point of Kalman iteration as suggested by Hammarberg (2002). If no data is present to make such an assumption, as a last resort a guess can be made regarding the state estimates and the uncertainty should be reflected in its error covariance matrix. A large variance, for example, signifies a possible large difference between the true and estimated state.

Ideally the initial values of the state estimates should be set to the expectation (mean) of the estimate where

$$\hat{\mathbf{x}}_0 = E\left[\mathbf{x}_0\right] \tag{6.13}$$

and the error covariance matrix to the variance of the state estimate such that

$$\mathbf{P}_0 = \mathrm{Var}\left[\mathbf{x}_0\right], \tag{6.14}$$

resulting in an *unbiased* estimate (Chui and Chen, 1987).

The consistency of the state estimates should be checked at the beginning of the filtering process to identify any possible problems. An inconsistency may be identified by large differences between the true and estimated states, which seem to increase over a finite amount of time (local divergence) or over all time (divergence). An inconsistency can reveal that the filter is biased or that the covariance matrix does not represent the true statistics of the initial estimate. Inconsistencies could also point to an inaccuracy between the real system and the model or that uncertainties specified by the process and measurement noise models are too modest.

Other factors affecting the consistency of the state estimates produced by the Kalman filter are observability, controllability and stability. The process noise driving the entire process, $\boldsymbol{\omega}_k$, is the "stochastic control" input, which serves to force the system to behave in a desired

manner. Its ability to do so correctly is determined by the *controllability* of the system, which is discussed in the following section.

6.6.4 Observability, Controllability and Stability

6.6.4.1 Introduction

Observability, controllability and stability are essentially derived from control theory but parallels or *dualities* exist between the control theory and estimation theory. While the knowledge of control theory facilitates the understanding surrounding issues in the estimation theory, it is not a prerequisite. See *Addendum A.6* for more details on observability, controllability and stability.

The significance of observability and controllability serves a main purpose in Kalman filtering theory, namely to observe and control the internal model in such a way as to best estimate the states through noisy measurements. The Kalman filter in effect needs to change its gain optimally in order to make the best estimate. It can however only do so by recursively passing the old estimates through a state transition matrix. It is thus dependent on the old estimate being able to affect the new estimate through all states (controllability) and these changes need to be observable through all states (observability).

6.6.4.2 Stability

Stability refers to the behaviour of state estimates when measurements are suppressed (Gelb, 1984). The stability of the Kalman filter is of great importance in that an unstable Kalman filter will result in divergence of the state estimate and thus render the estimator ineffective.

To obtain the unforced Kalman filter response (with suppressed measurements) the state estimation extrapolation (*Equation (6.11)*) needs to be substituted into the state estimation update equation (*Equation (6.9)*) resulting in

$$\hat{\mathbf{x}}_{k+1} = [\mathbf{\Phi}_k - \mathbf{K}_k\mathbf{H}_k]\,\hat{\mathbf{x}}_k. \tag{6.15}$$

For stability purposes it is desired that the estimates $\hat{\mathbf{x}}_{k+1}$ in *Equation (6.15)* tend to zero

as time progresses for any initial condition of $\hat{\mathbf{x}}_0$ (asymptotically stable).

Guaranteed asymptotical stability according to Gelb (1984) and uniqueness of the behaviour of $\mathbf{P}$ requires

1. stochastic uniform complete observability (*Equation (A.64)*),
2. stochastic uniform complete controllability (*Equation (A.65)*),
3. bounded $\mathbf{Q}$, $\mathbf{R}$ (from above and below) and
4. bounded $\mathbf{\Phi}$ (from above).

Complete observability and controllability are quite restrictive and in many practical cases of significance (Gelb, 1984) these conditions are not fulfilled. But in most cases Kalman filters designed in the usual way, as was discussed in *Section 6.6*, operate satisfactorily since *Equation (6.15)* frequently tends to zero over a finite time interval of interest, even though it might not be asymptotically stable in the strict sense. Instability is commonly associated with modelling errors and implementation considerations.

Since the Kalman filter has now been formally defined and issues such as initialisation, stability, controllability and observability addressed, the state-space equations can now be defined.

6.7 FORMATION OF STATE EQUATIONS

The state equations contain all state variables to be estimated, which cannot be measured directly. In the case of this auditory model the state equations should describe the following functional blocks shown in *Figure 6.4* (see *Subsection 4.4.4*).

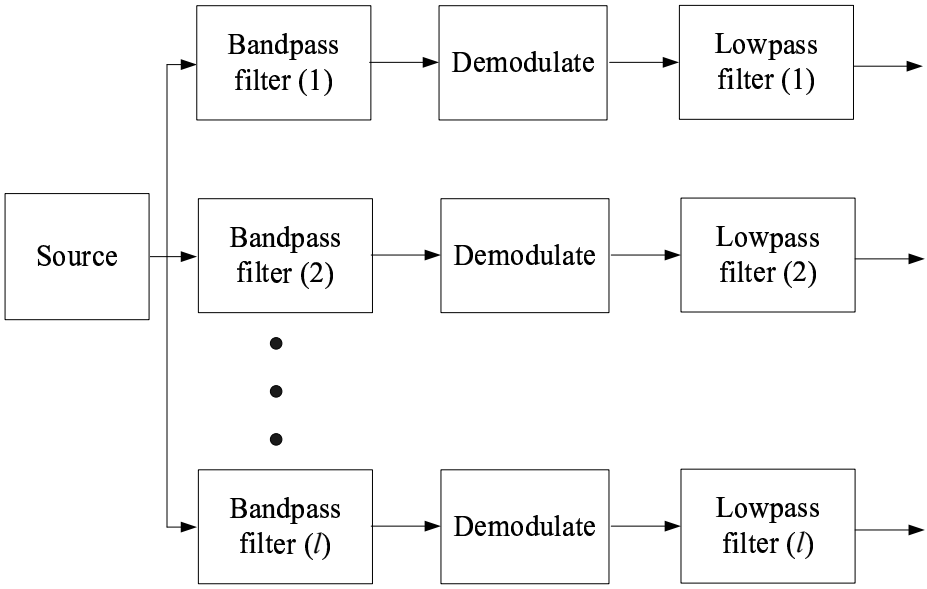

Figure 6.4: Signal decomposition block diagram.

Ideally each functional unit of *Figure 6.1* from the source signal to the coder/decoder pair should be represented by an equivalent state-space equation describing the functional unit. The coding/decoding and channel functional unit are not included in *Figure 6.4*. They will not be modelled as part of the state equations but rather as part of the measurement equation.

In order to model the coding/decoding and channel functional units as part of the state equations the discrete spike times generated by the coder would have to be modelled as separate state variables, which was the approach followed by Snyder (1975) in a related problem. Snyder defined the discrete spike times as a *doubly stochastic Poisson processes* that code an intensity function. The author then found a Kalman-Bucy filter to optimally decode the doubly stochastic Poisson process to estimate the intensity function. One problem with taking this approach in the current problem is that the spikes found on the noisy channel are assumed to be of Poisson nature, which is not necessarily true. Furthermore, it also does not separate the noise present on the channel from the information found in the spike train. Nevertheless, the linear decoder obtained in *Subsection 5.5.3* has a very similar impulse response

function to the one found by (Snyder, 1975).

In this research optimal decoding of the spike-train is avoided and decoding errors in conjunction with the channel noise are grouped as part of the measurement error of the power profile. For decoding purposes a linear decoder was used with transient response parameters closely related to the physiology as discussed in *Section 5.5*. The spike-decoding algorithm does not form part of the estimator implementation and decoding errors made by the spike decoder were assumed to be Gaussian noise. Even though the Gaussian noise assumption is a crude approximation of noise found on the channel and errors due to linear decoding of a non-linearly coded variable, it prevents the internal model of the estimator from becoming too complicated. The advantage of grouping the estimation error together with the channel noise in the measurement noise process of the Kalman filter, is that it allows for a modular design approach. Each different coding/decoding strategy can be compensated for by a different measurement noise model, resulting in a more generic and robust estimator.

6.7.1 State Equations for the Source Model

The first task in setting up the state-space equations was to find a suitable model for the source. As discussed in *Section 4.3* finding a suitable stochastic model for the speech model is not a trivial task. While unvoiced sounds are characterised by randomness, high-frequency content and relatively low amplitude, voiced sounds are more pulse like and quasi-periodic. While most designers refrain from making the speech signal model too complex (Moore, 1972) and often use a simple first-order lowpass filter (*Equation (4.1)*) driven by white-noise as a speech model, this oversimplifies the problem. The autocorrelation functions for the three random processes (amplitude, frequency and phase) defined in *Equation (4.2)*, were derived in *Section 4.3*. The autocorrelation function for the voiced sounds was derived for the case where the random amplitude and frequency process were of a Gaussian nature and the random phase uniformly distributed, as shown in *Equation (4.18)* and reproduced for convenience in *Equation (6.16)*.

$$\psi(\tau) = \frac{1}{2}\sigma_{\mathbf{r}}^2 \; e^{-(\tau\sigma_{\mathbf{f}})^2/2}. \tag{6.16}$$

In the case where the random frequency was chosen to have a non-zero mean μ the autocorrelation function was found to be

$$\psi(\tau) = \frac{1}{2}\sigma_{\mathbf{r}}^2 \cos(2\pi\mu\tau)\ e^{-(\tau\sigma_{\mathbf{f}})^2/2}. \tag{6.17}$$

The two autocorrelation functions are however only theoretical approximations of the source signal (defined in *Section 4.4*) and do not take into account that the *simulated* source signal (*Section 4.6*) has a constant amplitude, frequency and phase over 20 ms intervals. Nevertheless, it will be shown that similarities exist between the simulated and theoretical source signal (*Equation (6.17)*) and that the design methodology of finding a stochastic random process for the theoretical source signal is very similar to finding a stochastic random process for the simulated signal. In *Subsection 6.7.2* a periodic random process was designed to match that of the theoretical signal and then a similar approach was used to find a suitable periodic random process for the simulated signal.

From the two autocorrelation functions (*Equation (6.16)* and *Equation (6.17)*) it is evident that a complex shaping filter is required to stochastically model the source defined in *Section 4.3*. A single stochastic model as shown in *Figure 6.4* was preferred since it required only a single white-noise source and shaping filter to model a source signal containing *three* random variables (amplitude, frequency and phase). A simple narrow-band noise model (defined in *Addendum A.7*) could be used as a stochastic model for the source signal model as the result of passing white-noise through a narrow-band filter results in a random variable of the form given by *Equation (4.2)* with a random amplitude and phase. Since it however does not have a random frequency, it can only be used as a signal model for a particular bandlimited channel and would require not one but l source signal models for the l channels.

Thus, to maintain a single-source, multiple-channel system configuration as depicted in *Figure 6.4* a stochastic model fitting the source model as a *whole* was required. This model will be derived in *Subsection 6.7.2*, but first it is necessary to digress for a moment, to discuss correlated noise processes and the integration thereof with the discrete-time Kalman filter. This is done in *Subsection 6.7.1.1* to *Subsection 6.7.1.2*.

6.7.1.1 Correlated noise

As discussed in *Section 6.6.1*, the nature of the input noise driving the discrete-time Kalman filter has to adhere to certain conditions, one of these requires it to be a *white-noise* processes. A random variable is said to be *white* if its samples in time are *uncorrelated* (Haykin, 1994; Brown and Hwang, 1997; Grewal and Andrews, 2001; Papoulis and Pillai, 2002). The random variable g with random sequence $g_1, g_2, \ldots, g_n$ is uncorrelated if

$$E\left[\left(g_k - E\left[g_k\right]\right)\left(g_j - E\left[g_j\right]\right)^T\right] = \mathbf{G}\left(k, j\right)\delta(k-j), \tag{6.18}$$

where j and k are samples within the sequence n, $\delta\left(k\right)$ is the Kronecker delta function and $\mathbf{G}(k, j)$ the autocovariance matrix of the random variable g. Finding a stochastic model, which is not of white Gaussian nature as required by the discrete-time Kalman filter of *Section 6.6*, requires a shaping filter to convert the PSD of white Gaussian noise to the required PSD. By augmenting the state equations to include a shaping filter, a process coloured in time (ie. samples are correlated), can be generated from a white Gaussian noise source.

In order to achieve this a shaping filter as shown in *Figure 6.5* needed to be designed, with additional state variables, which feature in the model but are hidden from the output of the system (Lewis, 1986; Brown and Hwang, 1997; Grewal and Andrews, 2001).

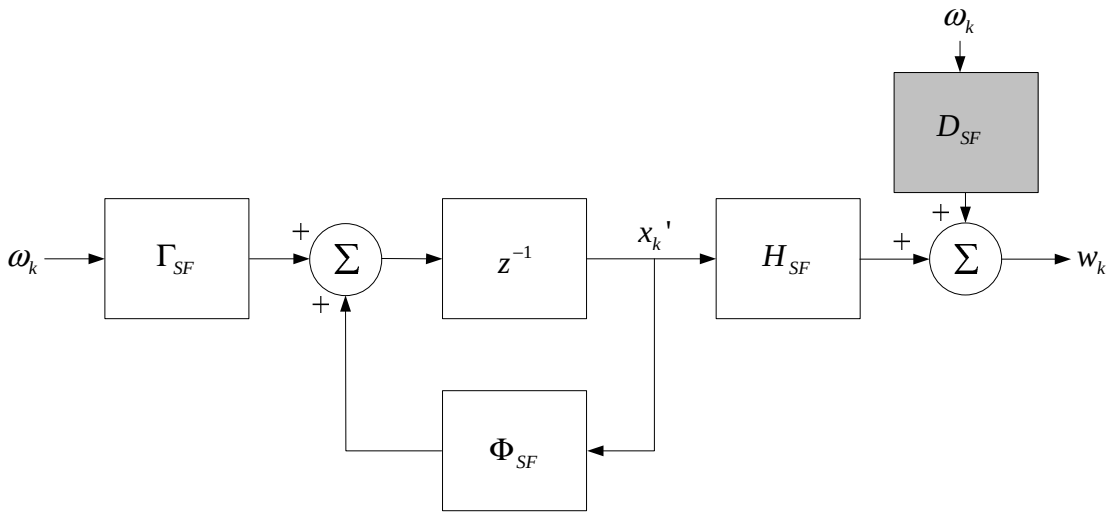

Figure 6.5: Discrete-time shaping filter model for non-white process noise w_k, generated from white Gaussian noise process ω_k.

A system of the form

$$\begin{aligned} x_{k+1} &= \Phi x_k + \Gamma \mathrm{w}_k \\ y_k &= H x_k + v_k, \end{aligned} \tag{6.19}$$

where $v_k \sim N(0, R)$, $x_0 \sim N(E(x)_0, P_0)$ (where $N(m, \sigma^2)$ is a Gaussian random variable with mean m and variance σ^2 and w_k is *not* white) and the PSD of the process w_k is given by $\Psi_W(f)$ that is not constant over all f (coloured), is a typical example where augmentation of the state vector is required.

If the autocorrelation function $\psi_W(z)$ is rational and if the PSD $|\Psi_W(z)| \neq 0$ for almost every z then there is a square, rational, asymptotically stable spectral factor $H(z)$ with zeros inside or on the unit circle such that

$$\Psi_W(z) = H(z)H^T(z^{-1}). \tag{6.20}$$

If $|\Psi_W(z)| \neq 0$ on $|z| = 1$ then $H(z)$ is strictly minimum-phase, i.e. all zeros are strictly inside $|z| = 1$. If *Equation (6.20)* can be spectrally factored, then the system $H(z)$ can be driven by white-noise $\omega_k \sim N(0, 1)$ to produce a spectral density $\Psi_W(z)$. The state realisation of $H(z)$ is given by

$$H(z) = H_{SF}(zI - \Phi_{SF})^{-1}\Gamma_{SF} + \mathrm{D}_{SF}, \tag{6.21}$$

where the coloured noise w_k is now given by the following spectrum shaping filter

$$\begin{aligned} x'_{k+1} &= \Phi_{SF}x'_k + \Gamma_{SF}\omega_k \\ \mathrm{w}_k &= H_{SF}x'_k + \mathrm{D}_{SF}\omega_k. \end{aligned} \tag{6.22}$$

Equation (6.22) is the *general* form of the shaping filter since it also contains a direct feed-through matrix D_{SF}. The direct feed-through has however been omitted since the state-space equations of the Kalman filter require this input for the measurement noise process. However, if a plant or system is of the form given by *Equation (6.19)* and an additional shaping filter is required to produce the required correlated process noise w_k, the direct feed-through can be 'absorbed' in the augmented process equation as will be discovered now.

Augmenting the state vector the process noise equation now becomes

$$\begin{bmatrix} x_{k+1} \\ x'_{k+1} \end{bmatrix} = \begin{bmatrix} \Phi & \Gamma H_{SF} \\ 0 & \Phi_{SF} \end{bmatrix} \begin{bmatrix} x_k \\ x'_k \end{bmatrix} + \begin{bmatrix} \Gamma \mathrm{D}_{SF} \\ \Gamma_{SF} \end{bmatrix} \omega_k, \tag{6.23}$$

with a measurement equation

$$y_k = \begin{bmatrix} H & 0 \end{bmatrix} \begin{bmatrix} x_k \\ x'_k \end{bmatrix} + v_k. \tag{6.24}$$

The direct feed-through matrix D_{SF} is now contained in the input matrix. If the shaping filter is however chosen to have $\mathrm{D}_{SF} = 0$ then the factor $\Gamma\mathrm{D}_{SF}$ simply reduces to zero. Shaping filters become useful tools when very specific and complex multi-stage colouring processes are desired as will be seen in *Section 6.6*.

The shaping filter can be integrated with the discrete-time Kalman filter of *Subsection 6.7.2* to provide for the case where the noise source ω_k is not white. In the next section it will be investigated how the simple Kalman filtering equations change when the filter is driven by a coloured noise process.

6.7.1.2 Discrete-time Kalman filter with coloured process noise

The classic discrete-time Kalman filter (Kalman, 1960) assumes that both the process and measurement noise are white Gaussian processes and that no cross-correlation exists between process and measurement noise. In reality however very few systems can be modelled as such and thus a method is required to convert or shape the non-white-noise sequence into the required form. In this case, the process noise is assumed to be correlated (coloured).

One way of accomplishing such a conversion is to rewrite the state equation in such a way so that the forcing function is driven by a zero mean, white Gaussian noise process as was discussed in the previous section. Given the standard discrete-time state equation

$$\mathbf{x}_{k+1} = \mathbf{\Phi}_k \mathbf{x}_k + \mathbf{\Gamma}_k \mathbf{w}_k \tag{6.25}$$

and measurement equation

$$\mathbf{z}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k, \tag{6.26}$$

where the process noise $\mathbf{w}_k$ is assumed to have correlated samples. The coloured process noise can however be written as a process driven by a zero-mean white Gaussian noise process $\boldsymbol{\beta}_k$. This conversion sacrifices complexity for a realistic source of process noise that is non-white where

$$\mathbf{w}_k = \mathbf{M}_{k-1}\mathbf{w}_{k-1} + \boldsymbol{\beta}_k, \tag{6.27}$$

where $\mathbf{M}_k$ is a suitable state transition matrix for the shaping filter. To use the previously defined Kalman filter equations from *Section 6.6*, new state variables have to be derived with the help of the shaping filter methodology. Letting

$$\tilde{\mathbf{x}}_k = \begin{bmatrix} \mathbf{x}_k \\ \mathbf{w}_k \end{bmatrix}, \tag{6.28}$$

$$\tilde{\boldsymbol{\Phi}}_k = \begin{bmatrix} \boldsymbol{\Phi}_k & 1 \\ 0 & \mathbf{M}_k \end{bmatrix} \tag{6.29}$$

and

$$\tilde{\boldsymbol{\beta}}_k = \begin{bmatrix} 0 \\ \boldsymbol{\beta}_k \end{bmatrix}, \tag{6.30}$$

the state equation can be rewritten as

$$\tilde{\mathbf{x}}_{k+1} = \tilde{\boldsymbol{\Phi}}_k\tilde{\mathbf{x}}_k + \tilde{\boldsymbol{\beta}}_{k+1}. \tag{6.31}$$

The measurement sensitivity matrix $\mathbf{H}_k$ also has to be adapted to include the second state variable $\mathbf{w}_k$ and can now be written as

$$\tilde{\mathbf{H}}_k = \begin{bmatrix} \mathbf{H}_k & 0 \end{bmatrix}. \tag{6.32}$$

Given *Equation (6.32)* the measurement equation now becomes

$$\mathbf{z}_k = \tilde{\mathbf{H}}_k \tilde{\mathbf{x}}_k + \mathbf{v}_k. \tag{6.33}$$

Comparing *Equation (6.32)* and *Equation (6.33)* with *Equation (6.3)* and *Equation (6.4)* and assuming the deterministic input $\mathbf{u}_k$ to be zero it can be seen that the augmented state and measurement equation are in the same form as that of the original state and measurement equations. The state equations can also be easily adapted to include the deterministic input if this is required, but the form of the state and measurement equations will remain the same. The Kalman filtering equations in *Section 6.6* can thus be used unchanged.

Correct and consistent initialisation

For correct and consistent initialisation the same guidelines hold as for the discrete-time Kalman filter in *Section 6.6*. According to Chui and Chen (1987) the initial state estimates should be set to

$$\hat{\mathbf{x}}_0 = \begin{bmatrix} E[\mathbf{x}_0] \\ 0 \end{bmatrix}, \tag{6.34}$$

since it was assumed that $\mathbf{w}_{-1} = 0$. Thus no extra information can be obtained from the past to make the best estimate and the best *unbiased* estimate of $\mathbf{w}$ thus becomes zero.

The initial error covariance matrix should be set such that

$$\begin{aligned}
\mathbf{P}_0 &= \text{Var}\left[\tilde{\mathbf{x}}_0 - \hat{\tilde{\mathbf{x}}}_0\right] \\
&= \text{Var}\left(\begin{bmatrix} \mathbf{x}_0 - E[\mathbf{x}_0] \\ \mathbf{w}_0 \end{bmatrix} + \right. \\
&\quad \left. \begin{bmatrix} \text{Var}\left[\mathbf{x}_0\mathbf{H}_0^T\right] \\ 0 \end{bmatrix} \begin{bmatrix} \mathbf{H}_0\text{Var}\left[\mathbf{x}_0\right]\mathbf{H}_0^T + R_0 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{H}_0 E[\mathbf{x}_0] - \mathbf{z}_0 \end{bmatrix}\right) \\
&= \begin{bmatrix} \text{Var}\left[\mathbf{x_0}\right] - \text{Var}\left[\mathbf{x}_0\right]\mathbf{H}_0^T\left[\mathbf{H}_0\text{Var}\left[\mathbf{x}_0\right]\mathbf{H}_0^T + R_0\right]^{-1}\mathbf{H}_0\text{Var}\left[\mathbf{x}_0\right] & 0 \\ 0 & \mathbf{Q_0} \end{bmatrix}
\end{aligned} \tag{6.35}$$

Since the shaping filter has been defined and the state-space equations derived a suitable *stochastic* model for the source signal can now be derived. A vital clue to which statistical model fits the source signal model the best is the requirement for a periodic autocorrelation function (see *Equation (6.17)*). As will become evident in the next section, a special form of the well-known Gauss-Markov (Gelb, 1984; Lewis, 1986) process provides just that.

Having discussed correlated noise processes and the integration thereof with the Kalman filter, a stochastic model for the source signal model in *Equation (6.17)* can now be found.

6.7.2 Stochastic Modelling of the Source Signal

From *Equation (6.17)* it can be seen that an exponentially decreasing periodic autocorrelation function is required to model the theoretical source signal (where $t_{skip} = 0$ s) defined in *Chapter 4*. Since the autocorrelation function and PSD are related to each other by the Fourier transform, stochastic modelling can be done either by matching autocorrelation functions or PSDs. Since the PSD of a random variable describes the random variable much more visually than the autocorrelation function, it was opted to match the PSDs rather than the autocorrelation functions. First the PSD of *Equation (6.17)* was derived. The PSD of a random variable can be obtained by taking the Fourier transform of the autocorrelation function such that

$$PSD(\omega) = \frac{1}{2}\sigma_{\mathbf{r}}^2 \int_{-\infty}^{\infty} \cos(2\pi\mu\tau)\ e^{-(\sigma_{\mathbf{f}}^2/2)\tau^2}\ e^{-j\omega\tau}\ d\tau. \tag{6.36}$$

By noting that a multiplication in the time domain results in a convolution in the frequency domain such that

$$g_1(t)g_2(t) \rightleftharpoons \int_{-\infty}^{\infty} G_1(\lambda)G_2(f-\lambda)\ d\lambda, \tag{6.37}$$

the autocorrelation function is divided into two separate functions g_1, g_2 and a constant term $\sigma_{\boldsymbol{r}}^2/2$. The variable $g_1(\tau)$ is then transformed such that

$$\begin{aligned} g_1(\tau) &= \cos(2\pi\mu\tau), \\ G_1(f) &= \frac{1}{2}\left[\delta(f-\mu) + \delta(f+\mu)\right] \end{aligned} \tag{6.38}$$

and then $g_2(\tau)$ such that

$$\begin{aligned} g_2(\tau) &= e^{-(\sigma_{\mathbf{f}}^2/2)\tau^2}, \\ G_2(f) &= \frac{\sqrt{2\pi}}{\sigma_{\mathbf{f}}}\ e^{-2(\pi/\sigma_{\mathbf{f}})^2 f^2}. \end{aligned} \tag{6.39}$$

Convolving *Equation (6.38)* with *Equation (6.39)* and scaling by the missing factor $\sigma_{\mathbf{r}}^2/2$ results in the following PSD (for $f > 0$)

$$PSD_{pos}(f) = \frac{\sqrt{\pi}}{\sigma_{\mathbf{f}}} \frac{\sigma_{\mathbf{r}}^2}{2\sqrt{2}} \, e^{-2(\pi/\sigma_{\mathbf{f}})^2 (f-\mu)^2}. \tag{6.40}$$

From *Equation (6.40)* it is evident that the PSD is an exponentially decaying function, located at the mean frequency μ. Since any random variable generated by passing a *zero-mean* Gaussian random variable through a linear system retains a zero-mean, generating a forcing function for the Kalman filter, which per definition has to have a zero mean, requires either a deterministic bias additional to that of a zero-mean forcing function or if possible a zero-mean forcing function with similar spectral properties such as that of the nonzero-mean random variable. A special form of the Gauss-Markov process, a zero-mean periodic random process, provides for such an interesting alternative to using a fixed bias and is evaluated next.

6.7.2.1 Gauss-Markov processes

A special class of random processes, generated by passing white-noise through linear time invariant filters, is the family of Gauss-Markov processes (Gelb, 1984), defined in *Addendum A.8*.

A first-order Gauss-Markov process, represented by the random variable X, has an auto-correlation function given by

$$\psi_X(\tau) = \sigma^2 \, e^{-\beta_1 |\tau|} + \mu^2, \tag{6.41}$$

with mean μ and PSD

$$\Psi_X(\omega) = \frac{2\beta_1 \sigma^2}{\omega^2 + {\beta_1}^2}. \tag{6.42}$$

A second-order Gauss-Markov process has an autocorrelation function given by

$$\psi_X(\tau) = \sigma^2(1 + \beta_2|\tau|)\ e^{-\beta_2|\tau|} + \mu^2, \tag{6.43}$$

with mean μ and PSD

$$\Psi_X(\omega) = \frac{4{\beta_2}^3\sigma^2}{(\omega^2 + {\beta_2}^2)^2}. \tag{6.44}$$

Comparing the Gauss-Markov processes with *Equation (6.16)*, similarities between the autocorrelation of the Gauss-Markov process and *Equation (6.16)* are apparent. Both autocorrelation functions decay exponentially and are multiplied by a variance. However, since a stochastic model for the random variable given by the autocorrelation function of *Equation (6.17)* is required, which contains a periodic term as well, a special form of Gauss-Markov process, the periodic random process, needs to be considered.

6.7.2.2 Periodic random processes

A random process X, which is closely related to the Gauss-Markov class of random variables is the periodic random process (PRP). While it is commonly listed as a second-order Gauss-Markov process (Lewis, 1986) it is in fact a special form in that it exhibits periodic behaviour. It is derived from the sinusoidal random process (Grewal and Andrews, 2001) given by autocorrelation function

$$\psi_X(\tau) = \sigma^2 \cos(\omega_n \tau) \tag{6.45}$$

and PSD

$$\Psi_X(\omega) = \pi\sigma^2 \left[\delta(\omega - \omega_n) + \delta(\omega + \omega_n)\right]. \tag{6.46}$$

In state-space form it can be written as

$$\dot{x}(t) = \begin{bmatrix} 0 & 1 \\ -{\omega_n}^2 & 0 \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \mathrm{w}(t). \tag{6.47}$$

The autocorrelation function of the sinusoidal random process can be slightly adjusted by adding an exponential term such that

$$\psi_X(\tau) = \sigma^2 \; e^{-\beta|\tau|} \cos(\omega_n \tau). \tag{6.48}$$

The autocorrelation function of the PRP contains both a periodic and exponential decaying function and the PSD of the PRP is given by

$$\Psi_X(\omega) = \frac{2\beta\sigma_{PRP}^2(\omega^2 + \alpha^2)}{\omega^4 + 2(2\beta^2 - \alpha^2)\omega^2 + \alpha^4}. \tag{6.49}$$

The differential equation of the periodic function is given by

$$\dot{x}(t) = \begin{bmatrix} 0 & 1 \\ -\alpha^2 & -2\beta \end{bmatrix} x(t) + \begin{bmatrix} 1 \\ (\alpha - 2\beta) \end{bmatrix} \mathrm{w}(t), \tag{6.50}$$

where

$$\alpha = (\beta^2 + \omega_n^2)^{0.5}. \tag{6.51}$$

Given the PSD of the PRP (*Equation (6.49)*) and that of the theoretical source signal model (*Equation (6.40)*) a close theoretical approximation can be found by adjusting the standard deviation of the periodic random process σ_{PRP}^2 and parameter β. Even though the similarities between *Equation (6.49)* and *Equation (6.40)* are not immediately evident it will be shown that close similarities between the two do exist.

The two PSDs of *Equation (6.49)* and *Equation (6.40)* were matched by selecting suitable values for σ_{PRP}^2 and β that minimised the mean-squared error between the two PSDs.

Table (6.3) summarises the values found.

Table 6.3: Statistics of theoretical source signal model (*Equation (6.40)*) and periodic random process (*Equation (6.49)*).

Model	Variable	Value
Theoretical source signal model (chosen)	$\sigma_{\mathbf{r}}^2$	1/6 (V^2)
	$\sigma_{\mathbf{f}}^2$	25 (Hz^2)
Periodic random process (matched)	β	4.63
	$\sigma_{\mathbf{PRP}}^2$	0.105

The resultant PSDs for the two random processes, the theoretical source signal and the PRP, are shown in *Figure 6.6*.

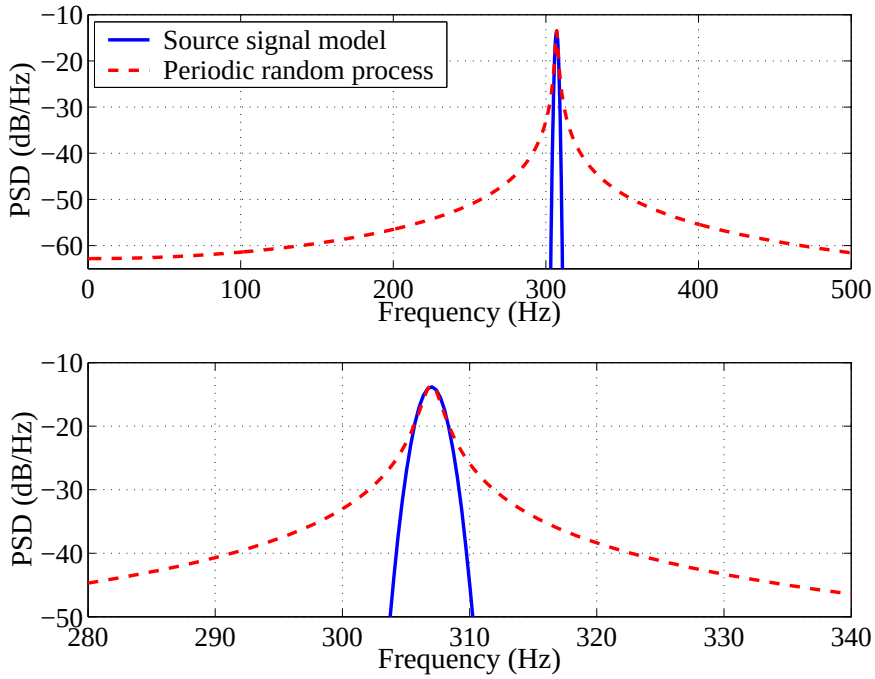

Figure 6.6: PSD comparison between theoretical source signal and PRP.

It is clear from *Figure 6.6* that indeed an equivalent periodic random process can be found to match the PSD of the theoretical source signal model. The PRP matches the PSD of the theoretical source signal model well in the 305 to 309 Hz range (around the mean value μ of 307 Hz) where almost all of the source signal model power is situated (-13 dB to -20 dB). The additional advantage of the PRP is that it has a zero mean and thus does not require a bias when implemented in the Kalman filter. The PRP PSD can also be easily adjusted, by

changing the variance σ^2_{PRP} and the parameter β, to match the PSD of a signal centred on any frequency μ Hz.

The PSD of the theoretical source signal differs from the simulated source signal (see results obtained in *Section 4.6*) since stationarity was assumed (*Section 4.3*, *Equation (4.9)*). However, since the simulated source signal is constant over a 20 ms interval, imitating the behaviour of a formant frequency in speech, the theoretical PSD of the theoretical source signal would be expected to become wider since the Fourier transform of a modulated rectangular pulse of 20 ms is a frequency shifted *sinc* function with a main-lobe width of $2/T$ where T is the period of the rectangular pulse. In this, case the main-lobe width is 100 Hz, which is relatively large compared to the 25 Hz frequency variance of the theoretical source signal. Thus, the PRP has to be matched to the simulated source signal. This will be done in *Subsection 6.7.3*.

However, the PRP discussed above is a continuous-time random process, which first needs to be discretised before it can be integrated with the discrete-time Kalman filter. It is thus necessary to digress for a moment to discuss the numerical evaluation of a continuous-time system.

6.7.2.3 Numerical evaluated, sampled continuous-time system

Given a continuous process

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t), \tag{6.52}$$

where $\mathbf{u}$ is a vector forcing function consisting of white-noise, the need often arises to sample such a process. Reasons can vary from having to use digital measuring devices, to needing to simulate a continuous-time process at discrete intervals in time. The sampling process can be done analytically or numerically where the former allows for an exact, closed-form solution and where the latter only provides an approximation. The numerical solution however avoids the time consuming task of inverse Laplace transforms, which become laborious for high-order systems as will be evident later on.

When sampling *Equation (6.52)* at discrete steps in time to obtain the difference equa-

tion relating the samples of $\mathbf{x}$, the solution at time t_{k+1} may be written as (Brown and Hwang, 1997)

$$\mathbf{x}(t_{k+1}) = \mathbf{\Phi}(t_{k+1,t_k})\mathbf{x}(t_k) + \int_{t_k}^{t_{k+1}} \mathbf{\Phi}(t_{k+1},\tau)\mathbf{B}(\tau)\mathbf{u}(\tau)\ d\tau, \tag{6.53}$$

or as

$$\mathbf{x}_{k+1} = \mathbf{\Phi}_k\mathbf{x}_k + \boldsymbol{\omega}_k, \tag{6.54}$$

where $\mathbf{\Phi}_k$ represents the state transition matrix from time t_k to t_{k+1} and $\boldsymbol{\omega}_k$ is the driven response at t_{k+1}.

The covariance matrix $\mathbf{Q}_k$ of discrete forcing function $\boldsymbol{\omega}_k$ can be found by letting

$$\begin{aligned} \mathbf{Q}_k &= E\left[\mathbf{x}_k\mathbf{x}_k^T\right] \\ &= \int_{t_k}^{t_{k+1}}\int_{t_k}^{t_{k+1}} \mathbf{\Phi}(t_{k+1},\varrho)\mathbf{B}(\varrho)E\left[\mathbf{u}(\varrho)\mathbf{u}^T(\varsigma)\right]\mathbf{B}^T(\varsigma)\mathbf{\Phi}^{\boldsymbol{T}}(t_{k+1},\varsigma)\ d\varrho\ d\varsigma. \end{aligned} \tag{6.55}$$

Given *Equation (6.52)* with initial conditions $\mathbf{x}(t_0) = \mathbf{x}_0$, the state transition matrix can be found by evaluating

$$\mathbf{\Phi}_k = \mathfrak{L}^{-1}\left[(s\mathbf{I} - \mathbf{A})^{-1}\right]|_{t=\Delta t}, \tag{6.56}$$

where the symbol $\mathfrak{L}^{-1}$ represents the inverse Laplace transform. For low-order systems *Equation (6.56)* gives a simple and straightforward, closed-form solutions. A similar situation applies to finding the discrete-time covariance matrix $\mathbf{Q}_k$. For higher-order systems the matrix-inverse however becomes quite laborious and at some stage virtually impossible to find a closed-form expression. It is thus sometimes easier to resort to a numerical method, especially if $\mathbf{A}$ is time-varying.

In the absence of a forcing function the state changes from t_k to t_{k+1}

$$\mathbf{x}(t) = \mathbf{\Phi}(t, t_k)\mathbf{x}(t_k), \qquad t \geq t_k, \tag{6.57}$$

which satisfies the matrix differential equation

$$\dot{\mathbf{\Phi}}(t, t_k) = \mathbf{A}(t)\mathbf{\Phi}(t, t_k), \qquad \mathbf{\Phi}(t_k, t_k) = \mathbf{I}. \tag{6.58}$$

If $\mathbf{A}$ is assumed to be constant over (t_k, t_{k+1}) then the state transition matrix is the matrix exponential given by

$$\mathbf{\Phi}_k = e^{\mathbf{A}\Delta t} = \mathbf{I} + \mathbf{A}\Delta t + \frac{(\mathbf{A}\Delta t)^2}{2!} + \ldots. \tag{6.59}$$

The numerical evaluation of $\mathbf{Q}_k$ is slightly more complicated and it has been shown by Brown and Hwang (1997) that $\mathbf{Q}_k$ must satisfy the differential equation

$$\begin{aligned} \dot{\mathbf{Q}}_k(t, t_k) &= \mathbf{A}(t)\mathbf{Q}_k(t, t_k) + \mathbf{Q}_k(t, t_k)\mathbf{A}^T(t) \\ &\quad + \mathbf{B}(t)\mathbf{W}\mathbf{B}^T(t) \qquad \mathbf{Q}_k(t_k, t_k) = 0, \end{aligned} \tag{6.60}$$

where $\mathbf{W}$ is the PSD matrix associated with the forcing function $\mathbf{u}$ in *Equation (6.52)*. Solving *Equation (6.60)* is no trivial matter and a method especially suited for **MATLAB** given by van Loan (1978) proceeds as follows:

First a $2n$ x $2n$ matrix is set up, where n is the number of state variables, such that

$$\mathbf{F}_k = \begin{bmatrix} -\mathbf{A} & \mathbf{B}\mathbf{W}\mathbf{B}^T \\ \mathbf{0} & \mathbf{A}^T \end{bmatrix} \Delta t. \tag{6.61}$$

Then solving for the matrix exponential (*expm* in MATLAB)

$$\mathbf{G}_k = e^{\mathbf{F}} = \begin{bmatrix} \cdots & \mathbf{\Phi}^{-1}\mathbf{Q}_k \\ \mathbf{0} & \mathbf{\Phi}^T \end{bmatrix}, \tag{6.62}$$

the solution for $\mathbf{\Phi}_k$ is obtained by taking the transpose of the lower-right partition of $\mathbf{G}_k$. Finally $\mathbf{Q}_k$ can be found by

$$\mathbf{Q}_k = \mathbf{\Phi}_k \cdot (\text{upper-right part of } \mathbf{G}_k). \tag{6.63}$$

Having discussed the discretisation process of any continuous-time state-space system, the principles can now be applied to discretise the continuous-time PRP given in *Subsection 6.7.2*. Subsequently suitable parameters for σ^2_{PRP} and the value β can be found, that match the PSD of the discretised periodic random process to that of the *simulated* source signal.

6.7.3 Simulated Periodic Random Process

It was shown in *Subsection 6.7.2.1* that the periodic random process with a zero mean and variance σ^2_{PRP} can be used to approximate the theoretical source signal model of *Equation (6.40)*. By finding suitable values for σ^2_{PRP} and β the periodic random process was adjusted in such a way as to emulate the first and second-order statistics of the theoretical source signal model. However, the assumption made was that the signal in *Equation (4.9)*, is wide sense stationary. The simulated source signal however imitates the behaviour of that of a formant frequency in speech as was mentioned earlier and defined in *Section 4.3*, where the amplitude, frequency and phase stays constant over a 20 ms interval after which the three values assume new random values. This process is similar to a modulated rectangular wave with random amplitude, frequency and phase. Since a modulated rectangular wave in time translates to a *sinc* signal in the frequency domain, where the main-lobe width is relatively large compared to the variance of the random frequency (100 Hz compared to 25 Hz) an increase in the value for parameter β in *Equation (6.51)* is expected. In order to find appropriate values for σ^2_{PRP} and β for the periodic random process it was chosen to minimise the mean-squared error between the *estimated* PSD (Yule-Walker spectral estimation method) of the simulated source signal and the theoretical PSD of the periodic random process.

With the aim of simulating the continuous-time periodic random process in discrete instants in times, a numerical evaluation of *Equation (6.50)* and $\mathbf{Q}(t)$, using the methods in *Subsection 6.7.2.3*, is undertaken. The differential equation of the periodic random process is given by *Equation (6.50)* and *Equation (6.51)* and needs to be discretised such that the state equation of the PRP can be written as

$$\boldsymbol{\beta}_{k+1} = \mathbf{L}_k \boldsymbol{\beta}_k + \boldsymbol{\omega}_k \tag{6.64}$$

and the measurement equation written as

$$\mathbf{z}_k = \mathbf{H}_k \boldsymbol{\beta}_k + \mathbf{v_k}. \tag{6.65}$$

$\mathbf{L}_k$ can be found by either solving *Equation (6.59)* or by taking the transpose of the lower-right partition of $\mathbf{G}_k$ in *Equation (6.62)*. The process noise covariance matrix $\mathbf{Q}_k$ can then be found by solving *Equation (6.63)*. For simulation and testing purposes the matrix $\mathbf{Q}_k$ is awkward to work with due to its non-diagonal nature (cross-correlation between two inputs). To avoid this it is preferable to find a matrix $\boldsymbol{\Upsilon}_k$ such that

$$\boldsymbol{\omega}_k = \boldsymbol{\Upsilon}_k \boldsymbol{\alpha}_k, \tag{6.66}$$

where $\boldsymbol{\alpha}_k \sim N(0,1)$. It is required that

$$E\left[(\boldsymbol{\Upsilon}_k \boldsymbol{\alpha}_k)(\boldsymbol{\Upsilon}_k \boldsymbol{\alpha}_k)^T\right] = E[\boldsymbol{\omega}_k \boldsymbol{\omega}_k^T] = \mathbf{Q}_k, \tag{6.67}$$

The matrix $\boldsymbol{\Upsilon}_k$ can be found by factorising it in such a way that

$$\boldsymbol{\Upsilon}_k \boldsymbol{\Upsilon}_k^T = \mathbf{Q}_k. \tag{6.68}$$

The elements of $\boldsymbol{\Upsilon}_k$ can be solved by *Cholesky factorisation* since $\boldsymbol{\Upsilon}_k$ is the *Cholesky factor* of $\mathbf{Q}_k$.

The numerically evaluated periodic random process of *Equation (6.50)* and *Equation (6.51)*

is then given by

$$\boldsymbol{\beta}_k = \mathbf{L}_{k-1}\boldsymbol{\beta}_{k-1} + \boldsymbol{\Upsilon}_k\boldsymbol{\alpha}_k, \tag{6.69}$$

where

$$\mathbf{L}_k = \begin{bmatrix} 0.88783 & 0.00023301 \\ -870.64 & 0.82946 \end{bmatrix} \tag{6.70}$$

and

$$\boldsymbol{\Upsilon}_k = \begin{bmatrix} 0.090971 & 0 \\ 85.239 & 32.151 \end{bmatrix}, \tag{6.71}$$

for all k.

The parameters found for the continuous-time PRP, given by *Equation (6.49)* to *Equation (6.51)* and the matched discrete-time PRP are summarised in *Table (6.4)*.

Table 6.4: Statistics of both simulated source signal model and simulated periodic random process (discrete).

Model	Variable	Value
Simulated source signal	$\sigma^2_{\mathbf{r}}$	1/6 (V^2)
model (chosen)	$\sigma^2_{\mathbf{f}}$	25 (Hz^2)
Discrete-time periodic	β	124
random process (matched)	$\sigma^2_{\mathbf{PRP}}$	0.100

The following simulated results were obtained for the periodic random process as discussed above. The periodic random process (*Equation (6.69)*) is shown at an arbitrary instant in time, in this case at 11 s into the simulation, in *Figure 6.7*.

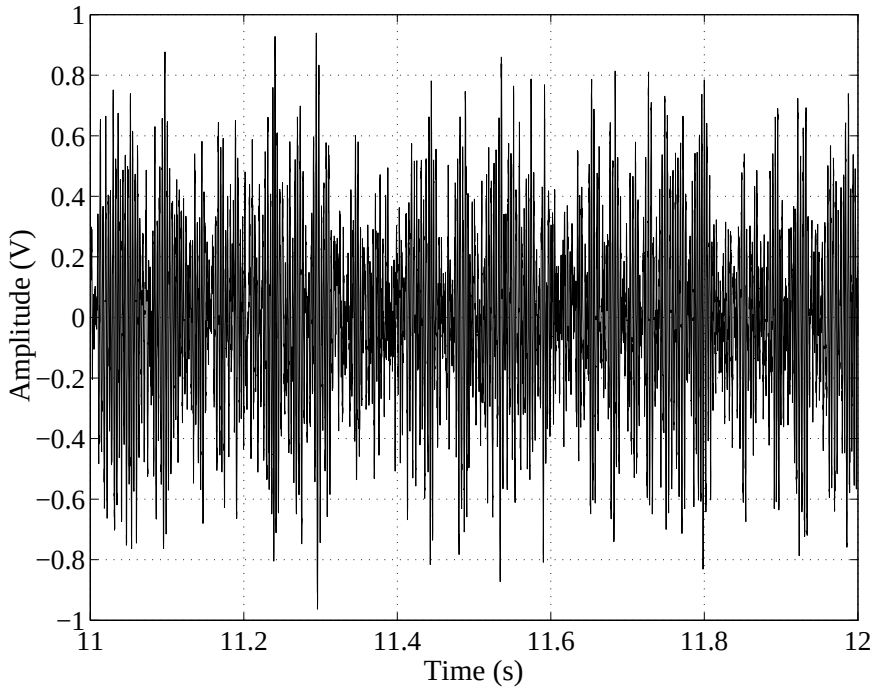

Figure 6.7: Simulated discrete-time periodic random process as a function of time.

There are obvious similarities between the periodic random process shown in *Figure 6.7* and the simulated source signal defined and simulated in *Chapter 4* and given by *Figure 4.9*. Both random signals have similar random amplitudes, situated between -1 V and 1 V. The obvious difference is that the simulated source signal has a constant amplitude, frequency and phase over an interval of 20 ms, while the periodic random process changes at every sample in time. Since the primary interest however lies in the PSD of the various signals, only the mean and variance are important in this case as far as the Kalman filter is concerned.

The PSDs for the simulated source signal, the theoretical periodic random process and the estimated PSD for the simulated PRP are shown in *Figure 6.8*.

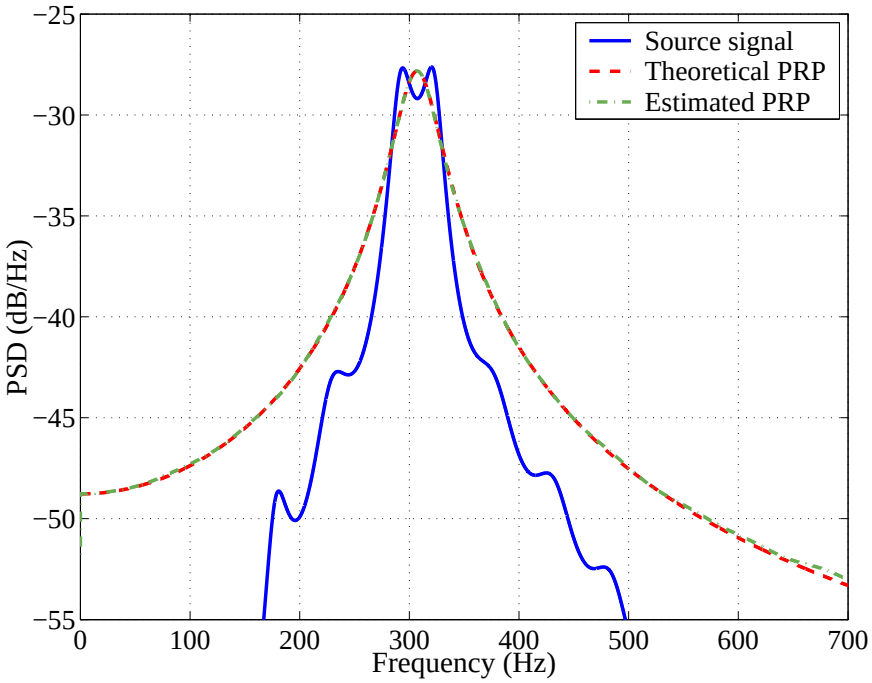

Figure 6.8: Estimated PSD of simulated source signal model, compared to theoretically obtained PSD of the periodic random process (PRP) and the estimated PSD of the simulated PRP. Estimated PSD of PRP obtained by means of Yule-Walker algorithm. PSD was estimated over 1 million random samples.

Care has to be taken when performing such a direct comparison, since the theoretical PSD of the PRP is an exact solution whereas the PSDs for the simulated source signal and the simulated PRP are *estimated* ones. The well-known Yule-Walker algorithm (Marple, 1987; Kay, 1988) was used to estimate the PSD from samples generated by the simulated source signal model and the simulated periodic random process. The PSD of the simulated source signal does not have a single distinct peak, but is (similar to the other two PSDs) situated around 305-307 Hz. The PSD (shape and magnitude) of the simulated periodic random process compares favourably to the theoretically obtained one and relatively well to the simulated source signal. The effect that the 20 ms interval has on the PSD of the simulated source signal is evident from *Figure 6.8* and the parameter β that was found (see *Table (6.4)*). As was expected the parameter β, determining the width of the PSD is much larger than was derived for the theoretical case given in *Table (6.3)*. Nevertheless, it is quite evident from the results that the simulated source signal, as far as the PSD is concerned, can be replaced by

a periodic random process with parameters given in *Table (6.4)*.

6.7.4 State Equations for Signal Extraction, Demodulation and Low-pass Filtering

From *Figure 6.4* the next functional unit that needs to be written in a set of system equations are the bandpass filters of the filter bank, where signal extraction for individual frequency bands takes place and then the power extraction by means of demodulation and low-pass filtering. Filters however can rarely be implemented directly in state-space form because of the stringent limitations set on them by the Kalman filter.

Finding an approximation of the actual filter can be done by means of spectral estimation algorithms such as the modified Yule-Walker algorithm, Prony's method, Linear Prediction, the Steiglitz-McBride method and inverse frequency design (These methods are available in MATLAB). In this case an ARMA model of desired order is found that models the response of the time-invariant linear filter as discussed in *Subsection 4.4.5*. Another method for approximating the filter would be to use the Kalman filter as a system identification tool and to estimate the characteristics of the bandpass filter (see Gelb (1984)).

In this case, however, the problem was not bound to a particular filter design and any filter, within limits, meeting the requirements set by the Kalman filter criterion could be designed. This was done in *Chapter 4*. An IIR filter bank was designed and an ARMA model was then used to implement the filter bank in the observable canonical form in *Subsection 4.4.5*. Similarly, a lowpass filter for each channel was implemented by using an ARMA model set up in the observable canonical form. Thus the state-space equations of the filters can be directly implemented in the system equations.

The process of demodulation is realized by squaring the signal at the output of the BPF. This results in a resultant DC component representing the sum of all frequency components present across the band, which provides an indication of the signal power in the spectral band. In order to obtain the resultant DC component on its own (removing the AC component) requires filtering of the result with a narrow-band lowpass filter, rejecting all other frequency components now situated at multiples of the fundamental frequency of the sampled signal.

This demodulation procedure is a *non-linear* operation, however a linear approximation is required for the Kalman filter. To investigate the implications of a non-linear functional unit on the Kalman filtering problem requires the understanding of the non-linear dynamics of the situation. It is thus necessary to digress for a moment to discuss some non-linear estimation theory and in particular a non-linear estimator know as the extended Kalman filter.

6.7.4.1 Non-linear estimation

Many dynamic systems are not linear and can be written in continuous-time as

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), t) + \mathbf{u}(t) \tag{6.72}$$

and

$$z(t) = \mathbf{h}(\mathbf{x}(t), t) + v(t), \tag{6.73}$$

or in discrete-time as

$$\mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}, k) + \boldsymbol{\omega}_k \tag{6.74}$$

and

$$z = \mathbf{h}(\mathbf{x}, k) + v_k, \tag{6.75}$$

where $\mathbf{f}$ and $\mathbf{h}$ are non-linear functions of $\mathbf{x}$. This provides potential problems since the Kalman filter is inherently a linear, optimal estimator. However, if it can be assumed that to some degree a dynamic system is linear over small perturbations, only slight alterations need to be made to the Kalman filter resulting in the *extended Kalman filter* (EKF). A general misconception is that the EKF applies methods of non-linear estimation theory to non-linear systems when in fact it applies the same methods of linear estimation theory (found in *Section 6.6*) to non-linear systems by assuming linearity over small perturbations. Nevertheless, it has been shown by Anderson and Moore (1979) that by means of ingenuity and appro-

priate simplifications, an EKF for non-linear dynamic systems can be found. An important distinguishing factor between the Kalman filter and the EKF is that the latter is *not* optimal (Anderson and Moore, 1979).

There are two main methods available to the design engineer to solve the problem of non-linearity: Linearisation about a *nominal* trajectory and linearisation about the *estimated* trajectory. The philosophy behind non-linear estimation is beyond the scope of this dissertation but interesting material can be found in Kolk and Lerman (1992) and Naumov (2000).

Non-linear Minimum Variance Estimation

In this research the focus will be on the *extended Kalman filter* (EKF), which is the most general non-linear minimum variance estimator. Two other methods for non-linear estimation are discussed briefly in *Addendum A.9* to give the reader an overview of the most important methods, each with their advantages and disadvantages and appropriate references. The EKF linearised about an estimated trajectory is discussed below.

Extended Kalman Filter

As was discussed, many dynamic systems are not linear yet the methodology of Kalman filtering can be applied to these problems with a slight variation in the Kalman filtering equations. The non-linear equations of *Equation (6.72)* and *Equation (6.73)* need to be linearised around an *estimated* trajectory (Grewal and Andrews, 2001; Brown and Hwang, 1997) in the case of the extended Kalman filter. Since only the sampled time domain is relevant here, the following discrete-time state and measurement equation are chosen:

$$\mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}, k) + \boldsymbol{\omega}_k, \tag{6.76}$$

$$\mathbf{z}_k = \mathbf{h}(\mathbf{x}, k) + \mathbf{v}_k. \tag{6.77}$$

The noise processes $\boldsymbol{\omega}_k$ and $\mathbf{v}_k$ are zero-mean white Gaussian noise processes with covariance matrices $\mathbf{Q}_k$ and $\mathbf{R}_k$.

Since an error exists between the true trajectory of $\mathbf{x}$ and the estimated trajectory $\hat{\mathbf{x}}$ such that

$$\mathbf{x}_k = \hat{\mathbf{x}}_k + \boldsymbol{\delta}\mathbf{x}_k, \tag{6.78}$$

where $\boldsymbol{\delta}\mathbf{x}_k$ is the error, *Equation (6.76)* and *Equation (6.77)* can be rewritten as

$$\hat{\mathbf{x}}_{k+1} + \boldsymbol{\delta}\mathbf{x}_{k+1} = \mathbf{f}(\hat{\mathbf{x}} + \boldsymbol{\delta}\mathbf{x}, k) + \boldsymbol{\omega}_k \tag{6.79}$$

and

$$\mathbf{z}_k = \mathbf{h}(\hat{\mathbf{x}} + \boldsymbol{\delta}\mathbf{x}, k) + \mathbf{v}_k. \tag{6.80}$$

The right-hand side of *Equation (6.79)* and *Equation (6.80)* can now be approximated by a Taylor series expansion of the form given in *Equation (A.61)*. By using only a first-order Taylor expansion approximation *Equation (6.79)* and *Equation (6.80)* can be rewritten as

$$\hat{\mathbf{x}}_{k+1} + \boldsymbol{\delta}\mathbf{x}_{k+1} \approx \mathbf{f}(\hat{\mathbf{x}}, k) + \left.\frac{\partial \mathbf{f}(\mathbf{x}, k)}{\partial \mathbf{x}}\right|_{\mathbf{x} = \hat{\mathbf{x}}_k(-)} (\mathbf{x}_k - \hat{\mathbf{x}}_k) + \boldsymbol{\omega}_k \tag{6.81}$$

and

$$\mathbf{z}_k \approx \mathbf{h}(\hat{\mathbf{x}}, k) + \left.\frac{\partial \mathbf{h}(\mathbf{x}, k)}{\partial \mathbf{x}}\right|_{\mathbf{x} = \hat{\mathbf{x}}_k(-)} (\mathbf{x}_k - \hat{\mathbf{x}}_k) + \mathbf{v}_k. \tag{6.82}$$

Since

$$\mathbf{x}_k - \hat{\mathbf{x}}_k = \boldsymbol{\delta}\mathbf{x}_k \tag{6.83}$$

and the fact that the unforced response can be defined as

$$\hat{\mathbf{x}}_{k+1} = \mathbf{f}(\mathbf{x}, k), \tag{6.84}$$

the linearised state-space equation and measurement equation can now be written as

$$\boldsymbol{\delta}\mathbf{x}_{k+1} \approx \left.\frac{\partial \mathbf{f}(\mathbf{x},k)}{\partial \mathbf{x}}\right|_{\mathbf{x}=\hat{\mathbf{x}}_k(-)} \boldsymbol{\delta}\mathbf{x}_k + \boldsymbol{\omega}_k \tag{6.85}$$

and

$$[\mathbf{z}_k - \mathbf{h}(\hat{\mathbf{x}},k)] \approx \left.\frac{\partial \mathbf{h}(\mathbf{x},k)}{\partial \mathbf{x}}\right|_{\mathbf{x}=\hat{\mathbf{x}}_k(-)} \boldsymbol{\delta}\mathbf{x}_k + \mathbf{v}_k. \tag{6.86}$$

The differential state transmission matrix $\boldsymbol{\Phi}^{[D]}$ is now formally defined and related to the non-linear function $\mathbf{f}(\mathbf{x},k)$ in the state equation such that

$$\boldsymbol{\Phi}^{[D]}(\hat{\mathbf{x}},k) = \left.\frac{\partial \mathbf{f}(\mathbf{x},k)}{\partial \mathbf{x}}\right|_{\mathbf{x}=\hat{\mathbf{x}}_k(-)}, \tag{6.87}$$

where

$$\boldsymbol{\Phi}^{[D]}(\hat{\mathbf{x}},k) = \left[\begin{array}{ccccc} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \frac{\partial f_1}{\partial x_3} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \frac{\partial f_2}{\partial x_3} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \frac{\partial f_3}{\partial x_1} & \frac{\partial f_3}{\partial x_2} & \frac{\partial f_3}{\partial x_3} & \cdots & \frac{\partial f_3}{\partial x_n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \frac{\partial f_n}{\partial x_3} & \cdots & \frac{\partial f_n}{\partial x_n} \end{array}\right]_{\mathbf{x}=\hat{\mathbf{x}}_k(-)}. \tag{6.88}$$

Similarly, the differential measurement sensitivity matrix $\mathbf{H}^{[D]}$, related to the non-linear function $\mathbf{h}(\mathbf{x},k)$ in the measurement equation by

$$\mathbf{H}^{[D]}(\hat{\mathbf{x}},k) = \left.\frac{\partial \mathbf{h}(\mathbf{x},k)}{\partial \mathbf{x}}\right|_{\mathbf{x}=\hat{\mathbf{x}}_k(-)}, \tag{6.89}$$

can be defined as

$$\mathbf{H}^{[D]}(\hat{\mathbf{x}},k)=\left[\begin{array}{ccccc}\frac{\partial h_1}{\partial x_1} & \frac{\partial h_1}{\partial x_2} & \frac{\partial h_1}{\partial x_3} & \cdots & \frac{\partial h_1}{\partial x_n}\\ \frac{\partial h_2}{\partial x_1} & \frac{\partial h_2}{\partial x_2} & \frac{\partial h_2}{\partial x_3} & \cdots & \frac{\partial h_2}{\partial x_n}\\ \frac{\partial h_3}{\partial x_1} & \frac{\partial h_3}{\partial x_2} & \frac{\partial h_3}{\partial x_3} & \cdots & \frac{\partial h_3}{\partial x_n}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ \frac{\partial h_n}{\partial x_1} & \frac{\partial h_n}{\partial x_2} & \frac{\partial h_n}{\partial x_3} & \cdots & \frac{\partial h_n}{\partial x_n}\end{array}\right]_{\mathbf{x}=\hat{\mathbf{x}}_k(-)}. \tag{6.90}$$

The EKF can be described by the following equations.

The Kalman gain matrix:

$$\mathbf{K}_k=\mathbf{P}_k(-)\mathbf{H}_k^{[D][T]}\left[\mathbf{H}_k^{[D]}\mathbf{P}_k(-)\mathbf{H}_k^{[D][T]}+\mathbf{R}_k\right]^{-1}. \tag{6.91}$$

The state estimation observable update:

$$\hat{\mathbf{x}}_k(+)=\hat{\mathbf{x}}_k(-)+\mathbf{K}_k\left[\mathbf{z}_k-\mathbf{h}(\hat{\mathbf{x}}_k(-))\right]. \tag{6.92}$$

The error covariance update:

$$\mathbf{P}_k(+)=\left[\mathbf{I}-\mathbf{K}_k\mathbf{H}_k^{[D]}\right]\mathbf{P}_k(-). \tag{6.93}$$

The error covariance extrapolation:

$$\mathbf{P}_{k+1}(-)=\mathbf{\Phi}_k^{[D]}\mathbf{P}_k(+)\mathbf{\Phi}_k^{[D][T]}+\mathbf{Q}_k. \tag{6.94}$$

The state estimation extrapolation:

$$\hat{\mathbf{x}}_{k+1}(-)=\mathbf{f}(\hat{\mathbf{x}}_k(+)). \tag{6.95}$$

Comparing the equations of the EKF to that of the Kalman filter in *Section 6.6* it can be seen how the estimators exactly differ from one another. The EKF (unlike the Kalman filter) is based on non-linear process and measurement model extrapolations (*Equation (6.95)*). However, the error covariance propagations are determined from the linearised state transmission matrix $\mathbf{\Phi}^{[D]}$ and measurement sensitivity matrix $\mathbf{H}^{[D]}$. Thus the state estimation update in *Equation (6.92)* is based on a non-linearly extrapolated state estimate but a linearly determined Kalman gain matrix.

The EKF (*Equation (6.85)*) can easily be related to the linear Kalman filter (*Equation (6.3)*) as will be shown next.

If $\mathbf{f}$ is a linear function of k such that

$$\hat{\mathbf{x}}_{k+1} = \mathbf{f}(\mathbf{x}, k) = \mathbf{\Phi}_k \mathbf{x}_k \tag{6.96}$$

and *Equations (6.78)* and *(6.96)* are substituted into *Equation (6.85)* then the linearised state-space equation reduces to the linear state-space equation suitable for the linear Kalman filter, where $\mathbf{\Phi}_k$ is some arbitrary state transition matrix. For this reason, mixing of linear and non-linear state-space and measurement equations, as will be done in *Subsection 6.7.5*, poses no problem, since no mathematical differences exist.

Correct and Consistent Initialisation

The same rules that apply to the Kalman filter in *Section 6.6* apply to the EKF with one exception. Additional problems related to the linearisation of the state transition matrix and measurement sensitivity matrix can lead to the divergence of the EKF. In this case, the designer should be careful in picking the sampling rate of the EKF. For more information see Brown and Hwang (1997) and Grewal and Andrews (2001).

Having defined the EKF and designed the accompanying stochastic models for each of the functional units in the previous sections the complete EKF state and measurement equations can now be setup.

6.7.5 Signal Extraction, Demodulation and Low-pass Filtering

Since the periodic random process has now been simulated, the entire EKF can be set up by concatenating the state equations of the functional units. The final state variable, the *power* found in a particular frequency band, or neural channel, can be written as

$$\epsilon_{k+1} = 0 \cdot \epsilon_k + \mathrm{w}_k. \tag{6.97}$$

The variable w_k is however not a white Gaussian noise process, but rather a coloured process, which will be defined in detail shortly. The measurement equation is of the form

$$z_k = H_k \epsilon_k + v_k, \tag{6.98}$$

where H_k is the measurement sensitivity matrix and v_k a white Gaussian measurement noise process with $N(0, \sigma^2)$. Since both the basic state and measurement equations have been found, the shaping filters for the periodic random process, bandpass filter bank, demodulator and LPF can be added. The state equation for the discrete-time periodic random process is given by *Equation (6.69)*. Appending these equations to the state equations of the bandpass filter bank the following two sequential shaping filters are obtained, namely that of the periodic random process given by

$$\boldsymbol{\beta}_{k+1} = \mathbf{L}_k \boldsymbol{\beta}_k + \boldsymbol{\Upsilon}_k \boldsymbol{\alpha}_k, \tag{6.99}$$

where $\boldsymbol{\Upsilon}_k \boldsymbol{\alpha}_k$ is as required a white Gaussian process with $N(0, \sigma^2)$ and the band-pass filtering process given by

$$\boldsymbol{\xi}_{k+1} = \mathbf{M}_k \boldsymbol{\xi}_k + \boldsymbol{\Omega}_k \boldsymbol{\beta}_k. \tag{6.100}$$

Demodulation is done by squaring the output of the bandpass filter bank, which is here discretised only after the appending process has been completed.

The next step is to append the state equation of the lowpass filter, which results in

$$\boldsymbol{\gamma}_{k+1} = \mathbf{N}_k \boldsymbol{\gamma}_k + \boldsymbol{\Xi}_k \left(\boldsymbol{\xi}_k^{\diamond}\right) \tag{6.101}$$

and

$$\mathrm{w}_k = \boldsymbol{\Gamma}_k \boldsymbol{\gamma}_k. \tag{6.102}$$

The diamond ($\diamond$) in *Equation (6.101)* denotes that the input into the lowpass filter is of a special form since the output of the bandpass filter needs to be squared before the DC component can be extracted from it. $\boldsymbol{\Xi}_k$ is thus a non-linear function of $\boldsymbol{\xi}_k$. The non-linear dynamics are however only discussed in the next section. The output of *Equation (6.102)* now contains the correctly coloured process noise w_k, which can drive the state variable of the power process ϵ_k. In order to show that *Equation (6.97)* to *Equation (6.102)* can be used with the discrete-time Kalman filter algorithm the state and measurement equations have to be augmented as was done in *Subsection 6.7.1.2*.

Define a new state variable vector that contains all state variable vectors as

$$\tilde{\mathbf{x}}_k = \begin{bmatrix} \epsilon_k \\ \boldsymbol{\gamma}_k \\ \boldsymbol{\xi}_k \\ \boldsymbol{\beta}_k \end{bmatrix}. \tag{6.103}$$

The dimensions of each of the state variables, representing a particular functional unit in the Kalman filter are summarised in *Table (6.5)* (single channel implementation).

Table 6.5: Dimensions of state variables of the functional units within the Kalman filter.

Functional Unit	**Symbol**	**Dimension**
Periodic random process	$\boldsymbol{\beta}$	(2 x 1)
Bandpass filter	$\boldsymbol{\xi}$	(7 x 1)
Lowpass filter	$\boldsymbol{\gamma}$	(4 x 1)
Power in channel	ϵ	(1 x 1)

The newly defined state transition matrix $\tilde{\mathbf{\Phi}}_k$ can be used to absorb all state transition matrices and process noise coupling equations of the preceding *Equations (6.99)* to *(6.102)*. By doing so the state transition matrix takes on the form

$$\tilde{\mathbf{\Phi}}_k = \begin{bmatrix} 0 & \mathbf{\Gamma}_k & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{N}_k & \mathbf{\Xi}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{M}_k & \mathbf{\Omega}_k \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{L}_k \end{bmatrix}, \tag{6.104}$$

where $\tilde{\mathbf{\Phi}}_k$ is a (14 x 14) matrix.

Combining the forcing functions so that it seems that the Kalman filter is driven by a white Gaussian noise function, results in

$$\tilde{\boldsymbol{\alpha}}_k = \begin{bmatrix} \omega_k^{\mathbf{1}} & 0 & 0 & 0 & \omega_k^{\mathbf{2}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{\Upsilon}_k \boldsymbol{\alpha}_k \end{bmatrix}^T, \tag{6.105}$$

where $\tilde{\boldsymbol{\alpha}}_k$ has dimensions (14 x 1). The purpose of adding the terms $\omega_k^{\mathbf{1}}$ and $\omega_k^{\mathbf{2}}$, both with dimensions (1 x 1), in *Equation (6.105)* will be discussed at the end of this section. For this moment in time, it will be assumed zero and the forcing function acting on the dynamic system influences the power state variable through the various dynamic systems mentioned earlier.

The augmented state equation can now be directly used by the Kalman filter algorithm and is given by

$$\tilde{\mathbf{x}}_{k+1} = \tilde{\mathbf{\Phi}}_k \tilde{\mathbf{x}}_k + \tilde{\boldsymbol{\alpha}}_k. \tag{6.106}$$

Since it was assumed that the channel noise could be modelled by simple white Gaussian noise, very little changes need to be applied to the measurement equation of *Equation (6.98)*. The assumption is made that none of the state variables $\boldsymbol{\gamma}_k$, $\boldsymbol{\xi}_k$ or $\boldsymbol{\beta}_k$ can be directly measured through the measurement sensitivity matrix making them "hidden" state variables. This is not necessarily a good assumption as far as Kalman filters are concerned since it requires that all state variables have to be visible through just one state variable, the power ϵ_k (i.e.

the observability issue as discussed in *Addendum A.6*). It also requires that all recursive changes applied by the process noise source $\tilde{\boldsymbol{\alpha}}_{k+1}$ can effect the output of the system (i.e. the controllability issue as discussed in *Addendum A.6*). From a biological point of view, making the state variables, other than the power ϵ_k, hidden ones, is a plausible one since no other sensory feedback mechanisms to the CANS exist in the auditory system before the point where the haircells innervate the auditory nerve bundle. In this case the augmented measurement sensitivity matrix is defined as

$$\tilde{\mathbf{H}}_k = \begin{bmatrix} H_k & 0 & 0 & 0 \end{bmatrix}. \tag{6.107}$$

The new measurement equation can then be written as

$$\mathbf{z}_k = \tilde{\mathbf{H}}_k \tilde{\mathbf{x}}_k + \mathbf{v}_k, \tag{6.108}$$

where $\mathbf{v}_k$ is defined as the measurement noise representing the channel noise from *Section 5.3*, absorbing both spike-jitter noise and the loss of spikes to unreliable retransmissions between synapses. As a last step before the state vector augmentation is complete the quadratic term in *Equation (6.101)* needs to be linearised. This is done next.

6.7.5.1 Including the non-linear dynamic process

Due to the non-linear demodulation process by means of squaring the signal at the output of the bandpass filter, modifications have to be made to the state-space equations derived above. The task is complicated by the fact that the non-linear term appears in the middle of the sequence of dynamic processes. However, by using the *observable canonical form* for the bandpass and lowpass filter and the fact that if

$$f(x,k) = A_k x_k, \tag{6.109}$$

then

$$\frac{\partial f(x,k)}{\partial x} = A_k, \tag{6.110}$$

the EKF equations reduce to the standard Kalman filter equations if the process is discrete and thus constant over a time interval. In this dissertation, only the quadratic term was linearly approximated, which requires careful implementation. The similarities between the Kalman filter algorithm and the EKF have to be exploited, which inherently are both *linear* and assume linearity when updating the error covariance and Kalman gain equations. One of the differences between the Kalman filter and the EKF is that the EKF assumes linearisation about the estimated trajectory, while the Kalman filter assumes that the discrete-time model is constant over a sample period. Another important difference is that the predicted state estimate and predicted measurement in case of the EKF are updated with non-linear functions f and h where

$$x_{k+1} = f(x,k) + \omega_k, \qquad \omega_k \sim N(0, Q_k) \tag{6.111}$$

and

$$z_k = h(x,k) + v_k, \qquad v_k \sim N(0, R_k). \tag{6.112}$$

The squared signal at the output of the bandpass filter passes directly into the lowpass filter and by choosing both the bandpass and lowpass filter to be in the discrete-time observable canonical form solving complex and time consuming matrix derivatives later are avoided, as will become evident shortly.

The measurement sensitivity matrix of the observable canonical form is of the form

$$\mathbf{C} = \begin{bmatrix} 0 & 0 & \dots & 0 & 1 \end{bmatrix}, \tag{6.113}$$

thus only the last element of the vector is passed through to the output. If the output of the bandpass filter thus has to be squared, only the last element is of concern. The squared output of the bandpass filter however has to drive the input of the lowpass filter as well and assuming that from *Equation (6.101)* the matrix $\mathbf{\Xi}_k$ has to absorb both the measurement sensitivity matrix and the coefficients of the zeros of the lowpass filter.

For a bandpass filter with n filter coefficients (number of pole-coefficients is n and number of

zero-coefficients is $n-1$)

$$\begin{aligned}\mathbf{f}_k(\boldsymbol{\xi}_k) &= \boldsymbol{\Xi}_k\left(\boldsymbol{\xi}_k^{\diamond}\right)\\ &= \begin{bmatrix}\Xi_{(1,k)}\\ \Xi_{(2,k)}\\ \vdots\\ \Xi_{(m,k)}\end{bmatrix}\left(\begin{bmatrix}0 & \dots & 0 & 1\end{bmatrix}\begin{bmatrix}\xi_{(1,k)}\\ \vdots\\ \xi_{(n,k)}\end{bmatrix}\right)^2\\ &= \begin{bmatrix}\Xi_{(1,k)}\xi_{(n,k)}^2\\ \vdots\\ \Xi_{(m,k)}\xi_{(n,k)}^2\end{bmatrix},\end{aligned} \tag{6.114}$$

where m is the number of filter coefficients of the lowpass filter (number of pole-coefficients is m and number of zero-coefficients is $m-1$).

To linearise *Equation (6.114)* the function $\mathbf{f}$ needs to be differentiated with respect to $\boldsymbol{\xi}$ such that

$$\frac{\partial \mathbf{f}_k}{\partial \boldsymbol{\xi}_k} = \boldsymbol{\Phi}_k^{[D]} = \begin{bmatrix}0 & & \boldsymbol{o} & 2\,\Xi_{(1,k)}\xi_{(n,k)}\\ & \ddots & & \\ \boldsymbol{o} & & 0 & 2\,\Xi_{(m,k)}\xi_{(n,k)}\end{bmatrix}, \tag{6.115}$$

where $\boldsymbol{o}$ represents the upper and lower triangular matrix, which in this case are zero. Finding the *a priori* and *a posteriori* error covariance matrix given by *Equation (6.93)* and *Equation (6.94)* requires the use of the linearised measurement sensitivity matrix and state transition matrix, which were just derived, whereas for the predicted state equation and predicted estimate on the measurement (*Equation (6.95)*, *Equation (6.92)*) the non-linear dynamic state and measurement equations have to be used.

It has to be noted here that the observable canonical form simplifies the matrix manipulations since for the controllable canonical form where the measurement sensitivity matrix is given by

$$\mathbf{C} = \begin{bmatrix} c_0 & c_1 & c_2 & \dots & c_{n-1} \end{bmatrix}, \tag{6.116}$$

Equation (6.114) would then be given by

$$\begin{aligned} \mathbf{f}_k(\boldsymbol{\xi}_k) &= \boldsymbol{\Xi}_k\left(\boldsymbol{\xi}_k^{\diamond}\right) \\ &= \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \left(\begin{bmatrix} c_{(1,k)} & c_{(2,k)} & \dots & c_{(m,k)} \end{bmatrix} \begin{bmatrix} \xi_{(1,k)} \\ \vdots \\ \xi_{(n,k)} \end{bmatrix} \right)^2 \\ &= \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \left(c_{(1,k)}\xi_{(1,k)} + c_{(2,k)} + \ldots + c_{(n,k)}\xi_{(n,k)}\right)^2 \end{bmatrix} \end{aligned} \tag{6.117}$$

and for notational convenience let

$$\boldsymbol{\upsilon}_k = c_{(1,k)}\xi_{(1,k)} + c_{(2,k)} + \ldots + c_{(n,k)}\xi_{(n,k)}, \tag{6.118}$$

then the derivative for *Equation (6.117)* is given by

$$\frac{\partial \mathbf{f}_k}{\partial \boldsymbol{\xi}_k} = \begin{bmatrix} 0 & 0 & \dots & 0 \\ & \vdots & & \\ 0 & 0 & \dots & 0 \\ 2\boldsymbol{\upsilon}_k c_{(1,k)} & 2\boldsymbol{\upsilon}_k c_{(2,k)} & \dots & 2\boldsymbol{\upsilon}_k c_{(n,k)} \end{bmatrix}, \tag{6.119}$$

which is significantly more intricate as far as implementation and computational complexity is concerned.

6.7.5.2 Robustness in tracking

Divergence of the state estimates from the true state variables is of great concern when designing a Kalman filter. Even if an exact representation of the real-world system to be tracked is found and implemented in the Kalman filter, convergence cannot be guaranteed. In fact, over longer periods of time during the tracking process the predicted error covariance matrix and hence the Kalman gain can tend to zero. The result is that the Kalman filter at some stage only believes its internal model and disregards any new measurements. Any new trends in the actual measurement data are thus not taken into consideration and divergence will occur. In order to avoid such an outcome, *fictitious* process noise has to be injected. This will prevent the Kalman filter from relying solely on its internal model, continuously readjusting its error covariance matrix and Kalman gain, never satisfied with the estimate, even if it happens to be equal to the true state variables. This is the effect that the term $\omega_k^{\mathbf{1}}$ in *Equation (6.105)* has. However, care has to be taken when choosing the magnitude of the fictitious noise, since it places a lower bound on the error covariance matrix (i.e. the Kalman filter bandwidth is limited from below).

Finding an exact model of a real-world system is a daunting task, which is further complicated if that system is non-linear. As was discussed in *Subsection 6.7.4.1*, linearisation about an estimated trajectory is employed by the EKF, resulting in possible approximation errors. Even though no errors as such are present in the internal model, approximations of the non-linear process can lead to divergence. The term $\omega_k^{\mathbf{2}}$ provides a technique to counteract divergence by allowing for some deviation from the internal model of the Kalman filter straight after the point of linearisation.

In this case two separate sources of fictitious process noise are added, one at the end of all the processes, to represent errors, which occur along the entire process and another after the linearisation procedure representing the error induced. In *Chapter 7* the advantages of this particular choice will become evident.

6.8 DISCUSSION AND CONCLUSION

In this chapter the state-space representations for each of the logical functional units designed in *Chapter 4* to *Chapter 5* were found. These were then implemented in the internal model of the EKF as was discussed at the beginning of the dissertation (see *Figure 1.5* reproduced for convenience in *Figure 6.9*).

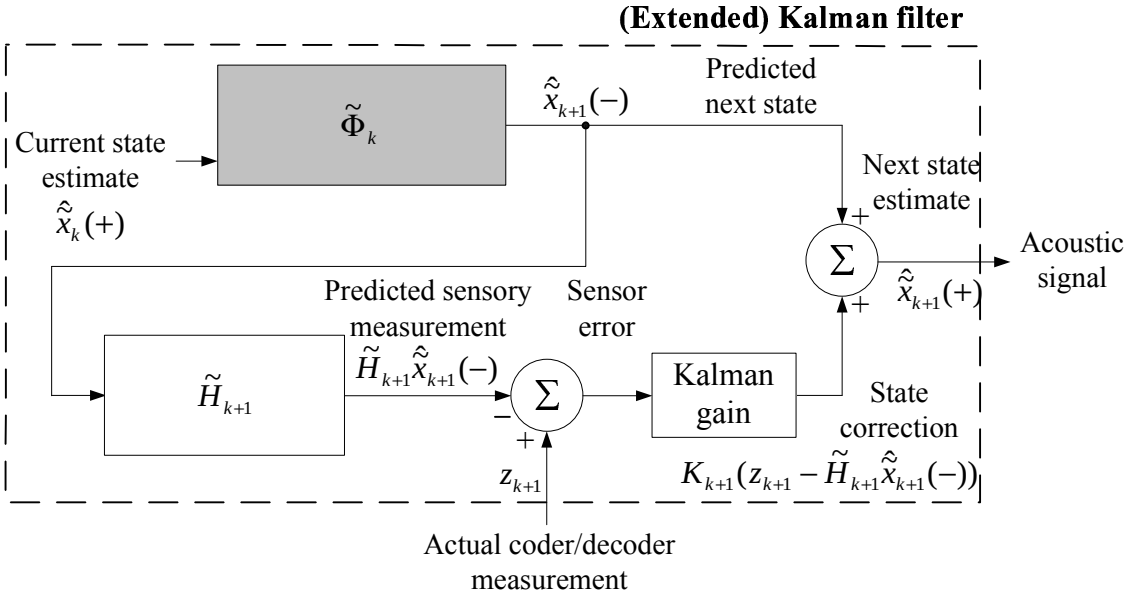

Figure 6.9: Kalman filter model for auditory system with appropriate symbols (modified from *Chapter 1.1*). $\tilde{\mathbf{\Phi}}_k$ is the augmented state transition matrix and $\tilde{\mathbf{H}}_k$ the augmented measurement sensitivity matrix.

The shaded block in *Figure 6.9* (Forward model of auditory system) is a stochastic representation of the auditory process defined by the source signal mode, the bank of bandpass filters, the demodulation functional unit and the lowpass filter shown in *Figure 6.2*. By means of state augmentation the various models were appended, which resulted in a state transition matrix of the form given by *Equation (6.104)*. The measurement sensitivity matrix of *Equation (6.107)* replaces the spike coder/decoder block in *Figure 1.5*. The actual measurement obtained by the auditory process model of *Figure 6.1*a is then compared to the estimated measurement of the internal model. Based on the error covariance matrix a suitable Kalman gain is found that minimises the MSE between the estimated state variables and the true state variables. Although the recursive estimation process is portrayed in *Figure 6.9*, the error covariance update procedure, based on the process noise and the measurement noise, is not, hence the exclusion of the noise variables $\mathbf{v}_k$ and $\tilde{\boldsymbol{\alpha}}_k$ found in *Equation (6.106)* and *Equation*

(6.108) from *Figure 6.9*.

The periodic random process from *Subsection 6.7.2* partially represents the shaded block in *Figure 6.9*. In this chapter it was shown that the statistics of the source signal as far as the mean and variance is concerned can be approximated by a periodic random process. The source signal can thus be replaced by a statistical model in the EKF, allowing for a better internal model, which in turn translates to a better estimate. Another advantage of using a periodic random process of zero mean is that it can model a source signal with just about any *non-zero* mean. Thus by adjusting the parameters σ^2_{PRP} and β of the periodic random process, any voiced sound model can be catered for.

In the next chapter the EKF derived in this chapter will be used to estimate the power signal, given the noisy measurements obtained from the auditory system model.

CHAPTER 7

EXTENDED KALMAN FILTER EVALUATION

7.1 AIM OF THIS CHAPTER

The main aim of this chapter is to test the formally stated hypotheses from *Chapter 1*, by examining how well the extended Kalman filter (EKF), containing a biologically plausible internal model, performs in the presence of decoding errors and spike channel noise specified in *Chapter 5*. The EKF is used to estimate the power process $x(t)$ generated by the auditory system given in *Figure 1.9* (repeated here as *Figure 7.1*). The implications of these findings are discussed and a conclusion is drawn.

7.2 INTRODUCTION

In the preceding chapters all the tools and models required for building the complete system were derived. As was discussed in *Section 1.2*, two separate systems were designed, the one simulating the behaviour of the envelope of a speech formant and the subsequent signal processing steps performed by the auditory system on this excitation signal as it passes through the auditory system and the second, an EKF containing a stochastically accurate internal model of the simulated auditory system.

The two separate systems first shown in *Figure 1.9* are reproduced below in *Figure 7.1*.

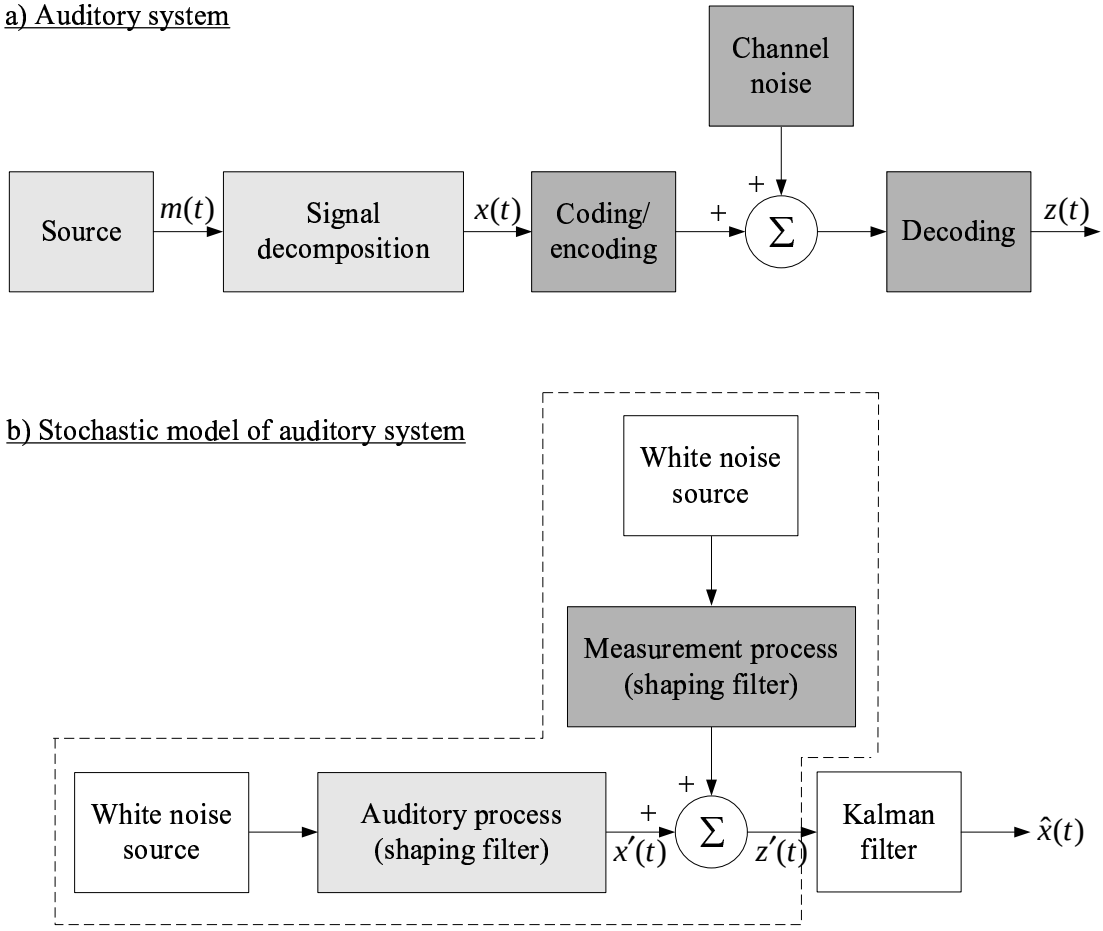

Figure 7.1: Auditory system model a) and stochastic model for a) shown in b). The Kalman filter contains an internal model of the stochastic model. $x'(t)$ and $z'(t)$ are the stochastic equivalents of $x(t)$ and $z(t)$.

The auditory system given in *Figure 7.1*a is used to generate the simulated measurement data $z(t)$ for the estimator model shown in *Figure 7.1*b, which in turn needs to estimate the state variables $x(t)$ of the auditory system. Note that the Kalman filter does not estimate the source signal $m(t)$. In this model, the source signal is assumed to be noiseless and the task of the Kalman filter is to compensate for internal noise, as explained in *Section 5.3*.

In the next section the entire auditory process that was designed in *Chapter 4* to *Chapter 5* is summarised briefly.

7.3 COMPLETE AUDITORY SYSTEM MODEL

The auditory system model as was discussed in *Section 1.2* can be modelled as a whole by *Figure 7.2*. This is a more detailed version of *Figure 7.1*a.

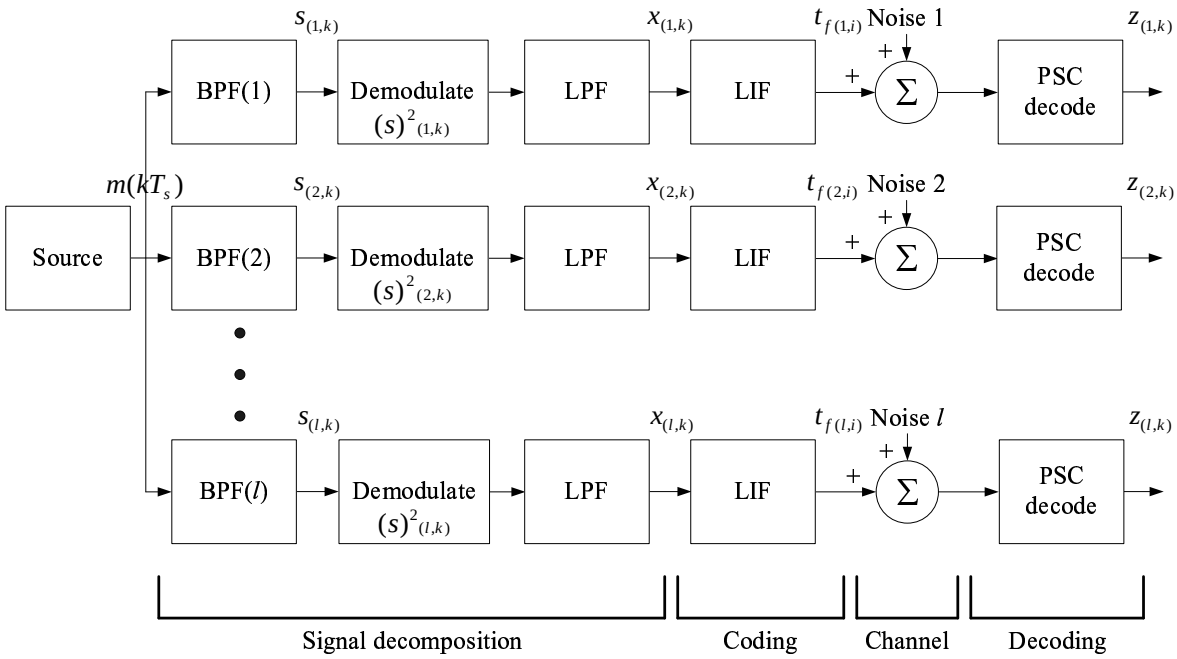

Figure 7.2: Complete auditory system model driven by randomly changing formant frequency (source signal), which in turn is decomposed into separate spectral bands from which the power is extracted, coded, transmitted over a noisy channel and then decoded. This is a discrete model with k the time index and T_s the sample period. The variables $x_{(l,k)}$ and $z_{(l,k)}$ correspond to $x(t)$ and $z(t)$ in *Figure 7.1*a, respectively.

The auditory system is driven by a source signal defined in *Section 4.3* with sample period T_s. It models the continually changing formant frequency of a source producing only voiced sounds. The envelope of the formant can change in amplitude, frequency and phase and its progress is tracked across the entire audible frequency range of roughly 20 Hz to 20 kHz by l evenly spaced but overlapping narrow-band bandpass filters, analogous to the functioning of a spectrum analyser. In order to determine the average power in a particular band each frequency component has to be demodulated to DC as was done in *Section 4.6*. The result is first low-pass filtered to remove unwanted frequency components before the true power is obtained. The power information is then coded by a LIF neuron pair allowing subsequent transmission over a noisy neural channel (see *Chapter 5*). The variable $\mathbf{t}_{f(l,i)}$ contains the spike times for both the "on" and "off" neurons on channel number l, with i the spike number.

For performance evaluation purposes, the signal is then linearly decoded providing a noisy measurement $z_{(l,k)}$ as required by the EKF estimator discussed in the next section.

7.4 EQUIVALENT STOCHASTIC MODEL (EKF INTERNAL MODEL)

Given the source signal and auditory system model in *Figure 7.2* the power $x_{(l,k)}$ present within each spectral band needs to be estimated. The EKF requires a stochastic equivalent model of the complete process in state-space form in order to make the best MMSE estimate. As was discussed in *Subsection 6.7.2* the implementation of a stochastic model, modelling a wide variety of different possible input sources, is a challenging task, especially if the inputs vary substantially over time. The best that can be done is to specify the source statistically in terms of its mean, variance and autocorrelation function over a long period in time. It is thus imperative to define a particular test signal and to find a statistical model that describes the signal best. In *Subsection 6.7.2* a stochastic equivalent model for the source was found and implemented.

This source model, the periodic random process, drives the spectral shaping blocks as is shown in *Figure 7.3*.

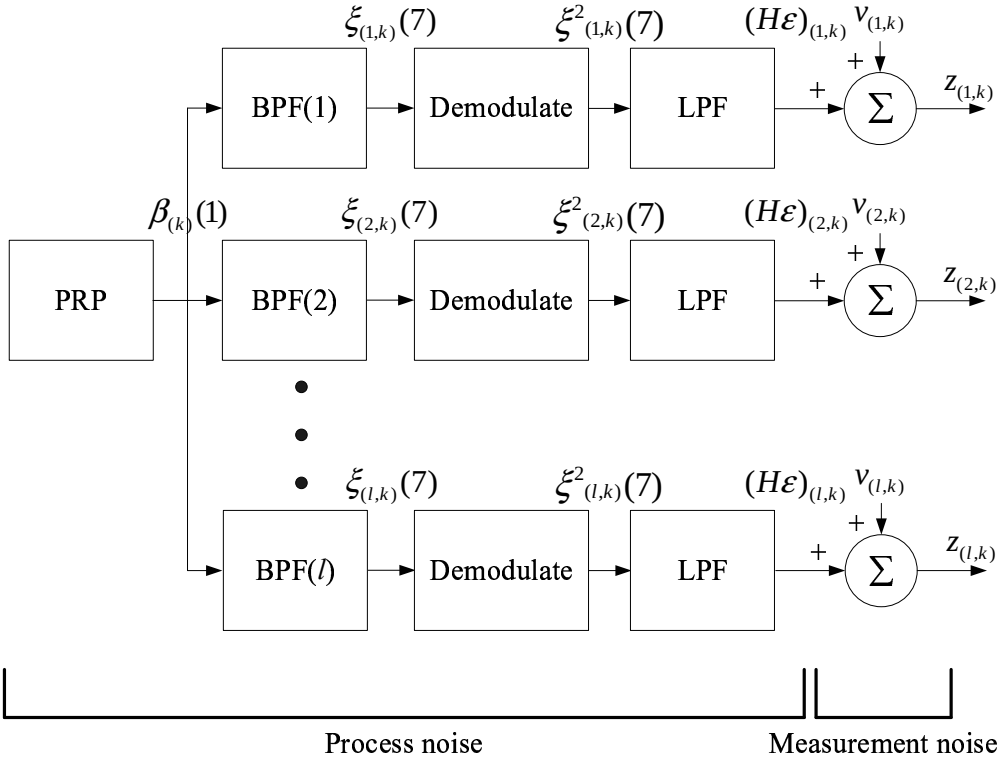

Figure 7.3: Equivalent stochastic model of the auditory process in *Figure 7.2*. The stochastic model consists of a process noise (state equations) and measurement noise process (measurement equations). The state equations are driven by one periodic random process (PRP).

Figure 7.3 shows the internal model of the EKF and the symbols used correspond to *Equation (6.97)* to *Equation (6.102)* in *Chapter 6*. The subscripts in brackets of the symbols in *Figure 7.3* represent the channel and sample number where the number in brackets behind the symbol indicates a specific element of the vector used, otherwise all elements are used. Since the source signal is decomposed by a bank of bandpass filters in the auditory system model, a set of equivalent shaping filters had to be found for the equivalent stochastic model. Seeing that the exact IIR filter coefficients are known due to the analysis-by-synthesis design approach, the filters are implemented directly in *Section 4.6* using an ARMA process.

The last part of the auditory system model that needs to be represented by an equivalent stochastic model is the coder, decoder and channel. This stochastic block was kept separate from the state equations and was implemented as measurement noise in the measurement equation of the EKF. This was done to avoid awkward doubly stochastic point processes

(Snyder, 1975) and simplified the EKF implementation substantially. The decoding mistakes made by the linear decoder on non-linear data (APs) combined with the unreliability of the channel was assumed to be Gaussian noise and implemented as Gaussian measurement noise.

The periodic random process drives the spectral shaping blocks representing the bandpass filter bank, demodulator and lowpass filter and can be represented by the augmented state equation (see *Subsection 6.7.5*) given by

$$\tilde{\mathbf{x}}_{k+1} = \tilde{\mathbf{\Phi}}_k \tilde{\mathbf{x}}_k + \tilde{\boldsymbol{\alpha}}_{k+1}. \tag{7.1}$$

Since only the state variables representing the power in each band are present, the measurement equation can be written as

$$\mathbf{z}_k = \tilde{\mathbf{H}}_k \tilde{\mathbf{x}}_k + \mathbf{v}_k. \tag{7.2}$$

In this dissertation the scope was limited to the design, implementation and simulation of a single channel ($l = 1$), since a multi-channel implementation in the EKF would require special attention. In the next section the EKF initialisation parameters are chosen.

7.5 EKF PARAMETER CHOICES AND INITIALISATION

In *Section 6.6* and *Subsection 6.7.1.1* the initialisation of the classic discrete-time Kalman filter and the Kalman filter with non-white process noise was discussed. The initialisation procedure is most often based on the statistics of the state variables at the beginning of the estimation process. Since the Kalman filter is based on white process and measurement noise, only the mean and variance is required to describe the state variables completely. It is however not incorrect to, pick a random sample out of a set of RVs described by a particular mean and variance for an initial estimate of the state variables. Generally a statistically initialisation approach is the best approach and is thus used in this dissertation.

For simulation purposes the entire system was initialised to zero initial conditions, which simplified the choice of initial state variables considerably. In this case it can be stated with certainty that the initial unbiased estimate is zero and the confidence in this estimate is

infinite (zero initial variance). The initial state vector is thus chosen as

$$\tilde{\mathbf{x}}_0 = \left[\begin{array}{c} \mathbf{0} \end{array} \right], \tag{7.3}$$

where $\tilde{\mathbf{x}}_0$ has dimensions (14 x 1). Statistically, the first state variables are chosen with zero error and thus a zero matrix for the initial error covariance matrix would seem apt. Assuming zero error in the initial estimate, however, allows the Kalman filter to rely solely on its internal model at the beginning of the estimation process, something that is never recommended unless it is certain that no error exists in either the initial estimates or the internal model. To avoid possible complications such as divergence or very slow convergence rates (locking) the error covariance matrix should be chosen very carefully.

The purpose of applying injection noise (see *Subsection 6.7.5*) after squaring (see *Figure 7.3*) and again after the calculation of power, allowed for some error or uncertainty. In this case it is wise to reflect the uncertainty of the estimate in the initial error covariance matrix. Special attention also has to be given to the periodic random process since the intermediate state variable was measured to have a variance of $0.4 \cdot 10^6$ V^2 (see *Subsection 6.7.3*). This state variable is a byproduct of the definition of the periodic random process and does not exist in the real model. Nevertheless, a statistical suitable initial mean and variance had to be chosen. The initial error covariance matrix for *all* simulations was chosen to be

$$\mathbf{P}_0 = \left[\begin{array}{cccccccccccc} 10^{-3} & 0 & 0 & 0 & 0 & 0 & & 0 & 0 & 0 & 0 \\ 0 & 10^{-9} & 0 & 0 & 0 & 0 & & 0 & 0 & 0 & 0 \\ 0 & 0 & 10^{-9} & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 10^{-9} & 0 & 0 & & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 10^{-3} & 0 & & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 10^{-9} & & 0 & 0 & 0 & 0 \\ \vdots & & & & & & \ddots & & & & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & & 10^{-9} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 10^{-3} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & & 0 & 0 & 0.1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & & 0 & 0 & 0 & 0.4 \cdot 10^6 \end{array} \right]. \tag{7.4}$$

No cross-correlation is assumed since all state variables were chosen to be the best, unbiased

estimates. The units for the diagonal of the matrix $\mathbf{P}_0(6:14, 6:14)$ are V^2 since the units for the random periodic process before and after the bandpass filter are measured in volts. After the demodulation process (squaring) the variance becomes V^4, represented by the diagonal of the matrix $\mathbf{P}_0(2:5, 2:5)$. The variance of the power (before conversion from power *units* to volts) is measured in V^4 since its units were formally defined to be in volts squared (see *Section 4.5*).

The process noise matrix $\mathbf{Q}_k$ was optimised for a particular driving process (the PRP) and model uncertainty (error in model and linearisation). The measurement noise matrix R_k on the other hand was optimised for a specific expected measurement noise. Thereafter the matrices were never changed since choosing a particular process and measurement noise is dependent on the real model, where changes have to be monitored by a higher hierarchial level observation system and appropriate matrices applied to the estimation problem. The motivation behind this assumption, is discussed in more detail in *Chapter 8*. The following process noise matrix was implemented.

$$\mathbf{Q}_k = \begin{bmatrix} 10^{-10} & 0 & 0 & 0 & 0 & 0 & & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 10^{-13} & 0 & & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & & 0 & 0 & 0 \\ \vdots & & & & & & \ddots & & & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0.090971 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & & 0 & 85.239 & 32.151 \end{bmatrix}, \tag{7.5}$$

where matrix location $\mathbf{Q}_k(1,1)$ represents the cumulative model uncertainty across the complete stochastic model and $\mathbf{Q}_k(5,5)$ the compensation factor for any linearisation errors (both tuned parameters). Besides the injection noise sources at matrix locations $\mathbf{Q}_k(1,1)$ and $\mathbf{Q}_k(5,5)$ only the variance of the driving process (the PRP) was required, which was defined and designed in *Chapter 6*. It was assumed that neither the cumulative model uncertainty nor the linearisation error are correlated with any other process since they account for uncertainties at *exact* positions within the equivalent stochastic model of *Figure 7.3* (also see *Subsection 6.7.5*) and correlations with all other points in the stochastic model are hard

to determine.

The measurement noise was optimised for an arbitrary spike-time jitter standard deviation of 1.5 ms and 4 % dropped spikes. This "tunes" the EKF for a particular amount of noise expected. Ideally this value should be estimated by another lower hierarchial level system and applied to the EKF by a higher hierarchial level observation system. Nevertheless, this arbitrary amount of measurement noise expected was sufficiently described by a white-noise process with variance given by

$$\mathrm{R}_k = 3.5 \qquad (\mathrm{V}^4). \tag{7.6}$$

The matrices for each of the stochastic functional units, derived in the previous chapters, are summarised here for convenience. The intermediate measurement sensitivity matrices for each of the stochastic functional units are also given. For the PRP of *Equation (6.99)* the input coupling matrix is given by

$$\boldsymbol{\Upsilon}_k = \begin{bmatrix} 0.090971 & 0 \\ 85.239 & 32.151 \end{bmatrix},$$

the state-transition matrix by

$$\mathbf{L}_k = \begin{bmatrix} 0.88783 & 0.00023301 \\ -870.64 & 0.82946 \end{bmatrix}$$

and the intermediate measurement sensitivity matrix (absorbed in the input coupling matrix of the following functional unit)

$$\mathbf{H}_k = \begin{bmatrix} 1 & 0 \end{bmatrix}. \tag{7.7}$$

In this case only the first of the two state variables of $\boldsymbol{\beta}_k$ is passed on to the bandpass filter.

In the case of the bandpass filter the input coupling matrix is given by

$$\mathbf{\Omega}_k = \left(\begin{bmatrix} -3.343 & 11.895 & -13.918 & 0.000 & 13.918 & -11.895 & 3.343 \end{bmatrix} \cdot 10^{-3} \right)^T, \tag{7.8}$$

the state-transition matrix by

$$\mathbf{M}_k = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & -0.946 \\ 0 & 1 & 0 & 0 & 0 & 0 & 5.105 \\ 0 & 0 & 1 & 0 & 0 & 0 & -12.073 \\ 0 & 0 & 0 & 1 & 0 & 0 & 15.904 \\ 0 & 0 & 0 & 0 & 1 & 0 & -12.297 \\ 0 & 0 & 0 & 0 & 0 & 1 & 5.2968 \end{bmatrix} \tag{7.9}$$

and the intermediate measurement sensitivity matrix by

$$\mathbf{H}_k = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{7.10}$$

In the case of the observable canonical form, only the last state variable is passed on to the lowpass filter. The process of demodulation was discussed in *Subsection 6.7.5* and is omitted here. In the case of the lowpass filter, the input coupling matrix is given by

$$\mathbf{N}_k = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0.996 \\ 0 & 1 & 0 & -2.991 \\ 0 & 0 & 1 & 2.996 \end{bmatrix}, \tag{7.11}$$

the state-transition matrix by

$$\mathbf{\Xi}_k = \left(\begin{bmatrix} 0.122 & -0.122 & -0.122 & 0.122 \end{bmatrix} \cdot 10^{-3} \right)^T, \tag{7.12}$$

and the intermediate measurement sensitivity matrix by

$$\mathbf{H}_k = \mathbf{\Gamma}_k = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}. \tag{7.13}$$

The augmented state equation (*excluding linearisation*, see *Subsection 6.7.5* for linearisation procedure) used by the EKF is

$$\tilde{\mathbf{x}}_{k+1} = \begin{bmatrix} \epsilon_{k+1} \\ \boldsymbol{\gamma}_{k+1} \\ \boldsymbol{\xi}_{k+1} \\ \boldsymbol{\beta}_{k+1} \end{bmatrix} = \begin{bmatrix} 0 & \mathbf{\Gamma}_k & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{N}_k & \mathbf{\Xi}_k^{\diamond} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{M}_k & \mathbf{\Omega}_k \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{L}_k \end{bmatrix} \begin{bmatrix} \epsilon_k \\ \boldsymbol{\gamma}_k \\ \boldsymbol{\xi}_k \\ \boldsymbol{\beta}_k \end{bmatrix} + \begin{bmatrix} \omega_k^{\mathbf{1}} \\ 0 \\ 0 \\ 0 \\ \omega_k^{\mathbf{2}} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \mathbf{\Upsilon}_k \boldsymbol{\alpha}_k \end{bmatrix}, \tag{7.14}$$

with measurement equation given by

$$\mathbf{z}_k = \tilde{\mathbf{H}}_k \tilde{\mathbf{x}}_k + \mathbf{v}_k, \tag{7.15}$$

where $\tilde{\mathbf{H}}_k$ is given by

$$\tilde{\mathbf{H}}_k = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{7.16}$$

The EKF was applied to the auditory model in *Figure 7.2*, its performance evaluated in the MSE sense and the results discussed in the next section.

7.6 FINAL RESULTS

In this section the final results obtained with the auditory and estimator system designed in the previous chapters are shown and discussed.

The EKF was used to estimate the power signal at $x^{'}(t)$ (see *Figure 7.1*b), i.e. just before the neural coding step. Since $x^{'}(t)$ is the stochastic equivalent of $x(t)$ in *Figure 7.1*a, it is already an estimate of the true power signal as found in a particular channel. The significance of the EKF as an estimator is however evaluated by how well it can estimate $x(t)$ in the presence of channel noise and decoding non-linearities. The EKF thus needs to be evaluated for different channel and decoding errors (as was specified in *Section 5.3*). This is done by determining the mean-squared error (MSE) between $x(t)$ and $\hat{x}(t)$ when a particular type of noise (spike-time jitter variance and spike loss) together with decoding errors due to non-linearities in the signal is present. However, before the EKF is evaluated in terms of its MSE performance, issues pertaining to observability, controllability and stability of the estimator are evaluated and discussed.

7.6.1 Observability, Controllability and Stability

Critical evaluation criteria of a Kalman filter are observability, controllability and stability. Even though the final behavioural results do not suggest instability, as the filter tracks the signal for the above chosen noise criteria, they do not guarantee that the Kalman filter will always be stable. In this case, evaluating the system in terms of observability, controllability are recommended.

First, each functional unit in the extended Kalman filter setup was evaluated independently by utilising the tools defined in *Subsection 6.6.4*. In this case it was sufficient to determine the observability and controllability matrices to prove that the system was observable and controllable since both $\mathbf{Q}_k$ (*Equation (7.5)*) and $\mathbf{R}_k$ (*Equation (7.6)*) were defined as positive definite matrices and the internal model was assumed to be stationary (i.e. constant system parameters as a function of time). In this case the observability and controllability grammians of *Equation (A.64)* and *Equation (A.65)* reduce to an algebraical expression in the form of *Equation (A.38)* and *Equation (A.39)*.

The result was that each process found in the internal model of the Kalman filter was determined to be completely observable and controllable since the observability and controllability matrices had full ranks. In case of the linearized process, the observability and controllability grammian satisfied the conditions given in *Addendum A.10.1* and *Addendum A.10.2* over the entire simulation interval.

7.6.2 Extended Kalman Filter Performance

Before the final results are shown and discussed, signal extracts illustrating how the EKF performs in the presence of noise, are depicted. Two scenarios for two different amounts of noise (both spike-time jitter and dropped spikes) are shown.

In *Figure 7.4* the influence that a change in noise intensity has on spike-timing is shown.

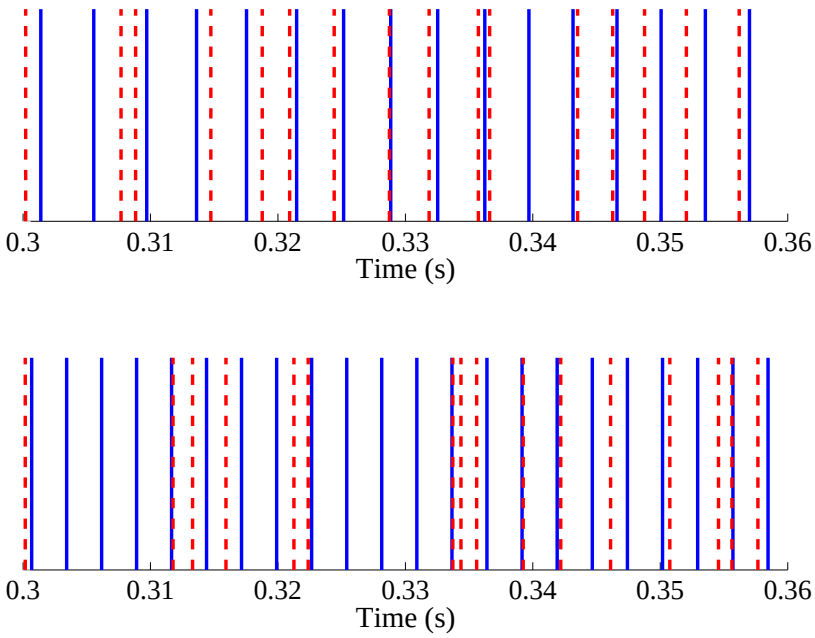

Figure 7.4: Corruption of original spike train (solid) by spike-time jitter and dropped spikes over an arbitrary interval in time. Small amounts of noise are added (spike-time jitter standard deviation of 1.5 ms and 4 % dropped spikes) to the spike train at the top of the figure while large amounts are added to the spike train at the bottom of the figure (spike-time jitter standard deviation of 2 ms and 15 % dropped spikes). New spike positions of corrupted spike train are indicated by dashed lines.

At the top of *Figure 7.4* a relatively small amount of noise is added compared to the spike train at the bottom of the figure. Both spike-time deviations and the loss of spikes are evident from *Figure 7.4*. In the first extract, spike-time jitter noise (standard deviation of 1.5 ms) was applied and spikes dropped randomly (4 %) on one of the l channels (bandpass filter centred around 300 Hz).

The first 1.5 s of the power in that particular channel are shown in *Figure 7.5*.

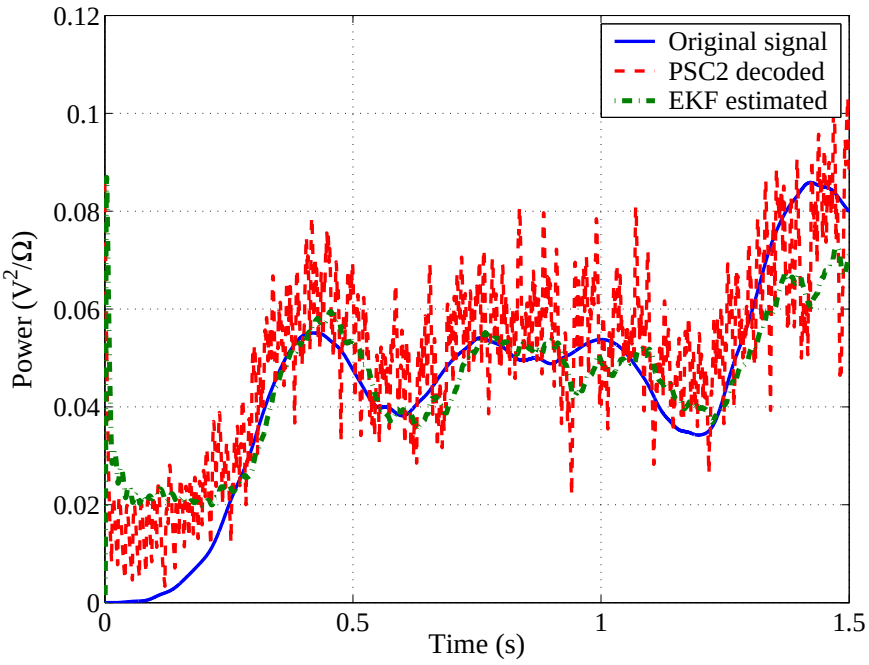

Figure 7.5: EKF performance in the presence of spike-time jitter noise with standard deviation of 1.5 ms and 4 % randomly dropped spikes. This is one of the l channels, with the original power signal being coded by a LIF neuron pair, which is subsequently transmitted across a noisy neural channel, linearly decoded by a PSC2 filter and then estimated by the EKF.

Before time zero, no power is present in the channel since no stimulus has been applied to the system yet. From the response of the EKF it is evident that the time it takes for "locking" onto the undistorted power signal (original signal) is minimal (< 0.185 s) since the initial value for the power signal is chosen to be zero, which is exactly from where the power process initially starts off. From *Figure 7.5* it is however also evident that even though the process is initialised from zero, the filter tends to be "unstable" for a short period of time, before locking eventually takes place.

The amplitude of the power signal is far smaller than one, since the amplitude of the formant frequencies is of a Gaussian nature with mean and variance given by $N(0, 1/6)$. A gain stage is used to allow operation in the active region of the neuron-pair as discussed in *Section 5.5*.

From *Figure 7.5* it is evident that the PSC2 decoder derived from the optimal decoder in *Section 5.4*, which is optimal only in the absence of noise, is corrupted by the channel noise. The EKF maintains the shape of the original signal well and "smooths" the spiky output of the PSC2 decoder. The EKF's sensitivity to measurements can be set by changing the measurement noise parameter in the EKF. However, larger than expected measurement noise forces the EKF to rely more on its internal model and its initial estimates (states and error covariance). This can have a detrimental effect on the estimates since any errors in the internal model of the EKF can lead to divergence of the estimate. If the approximate measurement noise is known, it provides for a suitable point of reference to start from. Additional tuning, however, is almost always required in addition to a guessed or approximated point of reference.

Comparing the difference in MSE between the PSC2 decoded signal relative to the original signal and the EKF estimated signal relative to the original signal, the EKF outperforms the PSC2 filter by an order of 3.4 (5.3 dB).

The following extract illustrates how the performance of the EKF changes when the channel noise is more severe (see also bottom of *Figure 7.4*).

In *Figure 7.6* spike-time jitter noise with a standard deviation of 2 ms is applied and 15 % of the spikes are dropped randomly.

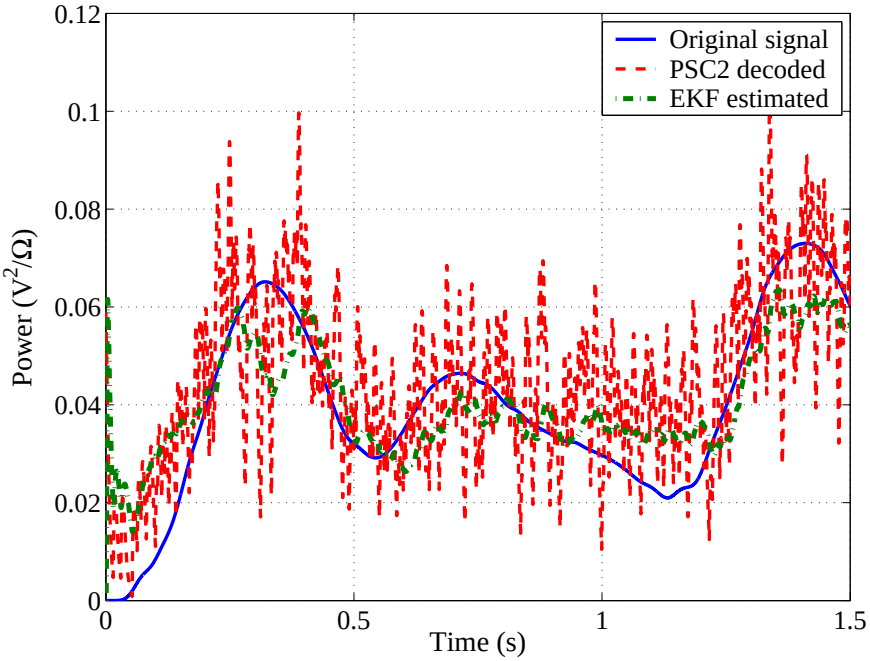

Figure 7.6: EKF performance in the presence of spike-jitter noise with standard deviation of 2 ms and 15 % spikes dropped randomly. This is one of the l channels, with the original power signal being coded by a LIF neuron pair, which is subsequently transmitted across a noisy neural channel, linearly decoded by a PSC2 filter and then estimated by the EKF.

A 2 ms spike-time jitter standard deviation corresponds to the gap that exists between spikes at a constant spike rate of 500 spikes per second, where 500 spikes per second is in this case the maximum spike rate.

The time it takes the EKF to lock onto the original signal, in *Figure 7.6*, is roughly 0.25 s. Since the variance of the measurement noise parameter in the EKF remains unchanged from that of the previous example, the estimator struggles to cope with a larger than expected measurement error variance. The reasoning behind a constant measurement noise variance is as follows. Even though the channel noise is expected to vary over time, the EKF on its own cannot detect these changes. In reality, in the case of the auditory system, it can be assumed that another lower hierarchial system estimates the noise in the channel, which a higher hierarchial level observatory system utilises to tune the measurement noise of the EKF. This way, the measurement noise could always be chosen optimally (in the case where another

Kalman filter estimates the noise), allowing for a better overall EKF performance.

From *Figure 7.6* it is evident that the PSC2 decoder, optimised for a noiseless spike train, does relatively poorly in the presence of the two noise processes. The EKF however still does well in tracking the amplitude and shape of the original signal.

The two power signal extracts shown above give an indication of how the underlying EKF performs under different conditions of noise. Using similar simulations the EKF is now evaluated over a range of values for the two noise processes (spike-time jitter and dropped spikes) found on the channel. To compare performances from different simulation runs, it has to be kept in mind that although the statistics of the source signal are constant, a new random sample for the amplitude, frequency and phase are chosen every 20 ms (see *Section 4.3*). The random power signal is thus a slowly changing random variable and care has to be taken not to limit the size of the ensemble of samples, which could lead to bad or inaccurate statistical inference in terms of the MSE.

The simulations were time-intensive and a balance between statistical precision and reasonable simulation times had to be found. A set of 20 runs of 4 s each, at a sampling rate of 4 kHz was chosen as a statistical set, which represents a pseudo-random sequence of 320 000 samples.

Since the initial time to "locking" was small and the settling time of the random process never greater than 0.5 s, the first 0.5 s of each of the 20 runs was rejected when determining the MSE. Thus the set contained 280 000 samples during which the amplitude, frequency and phase of the formant frequency changed 3 500 times. Once the set was defined, simulated and stored, an optimal decoding filter for the set (70 s) was found from which the parameters of the PSC2 filter were derived.

The final results are shown in *Figure 7.7* and *Figure 7.8*. These results were set out at the beginning of the dissertation as to be obtained.

First the MSE was evaluated as a function of spike-time jitter noise with a standard deviation ranging from 0.5 ms to 7.5 ms. No spikes were dropped and the results are shown in *Figure 7.7*.

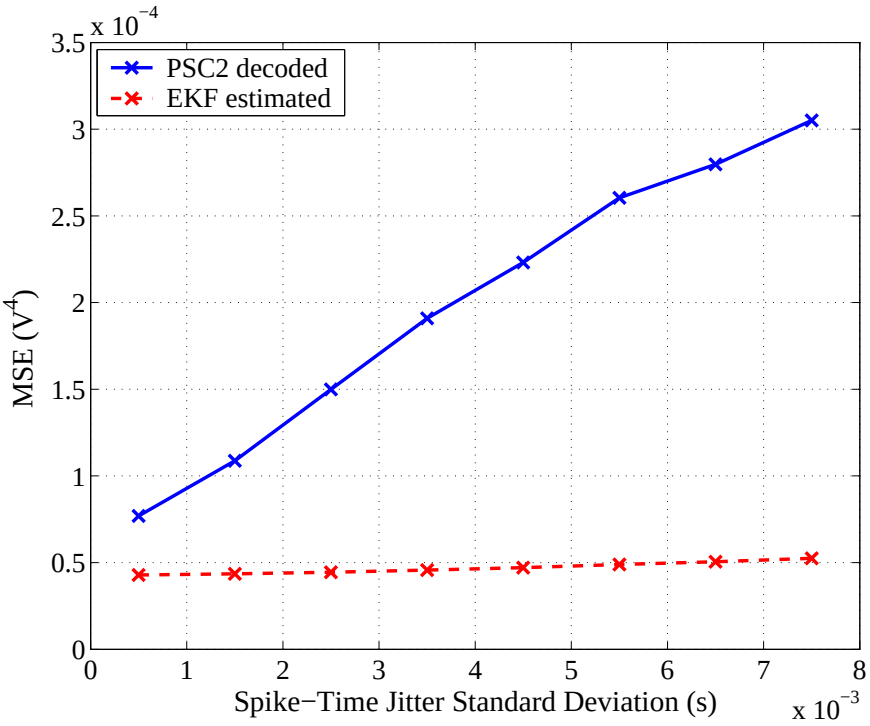

Figure 7.7: System performance in the presence of spike-time jitter noise for PSC2 decoded and EKF estimated signals.

As expected, the PSC2 filter performs at its best at small amounts of spike-time jitter noise as it was optimised to decode a noiseless spike train. Similarly, the EKF, tuned for low spike-time jitter noise, performs the best at small amounts of spike-time jitter noise. As expected, the MSE for both the PSC2 decoded signal and the EKF estimated signal increases with an increase in spike-time jitter variance, however, the EKF on average outperforms the PSC2 decoding filter. This can be attributed to the fact the EKF has prior knowledge of the underlying auditory system model and noise model while the PSC2 filter, a fixed filter, was tuned to decode noiseless spike trains. It is thus not surprising that the slope of the MSE versus spike-time jitter standard deviation is smaller for the EKF than the PSC2 decoding filter. Even though errors in the expected measurement noise model cause the Kalman filter to operate as a suboptimal estimator at spike-time jitter noise variances other than for what it was optimised, its *a priori* knowledge still makes it a more suitable estimator than a static filter such as the PSC2 filter. The minimal degradation in performance in the MMSE

error sense in the case of the EKF is however unexpected. The EKF performs exceptionally well compared to the PSC2 decoding filter, over the *entire* range of spike-time jitter variances.

In the second case, the MSE was evaluated as a function of the number of dropped spikes, which ranged from 0 % to 40 %. No spike-jitter noise was added and the results are shown in *Figure 7.8*.

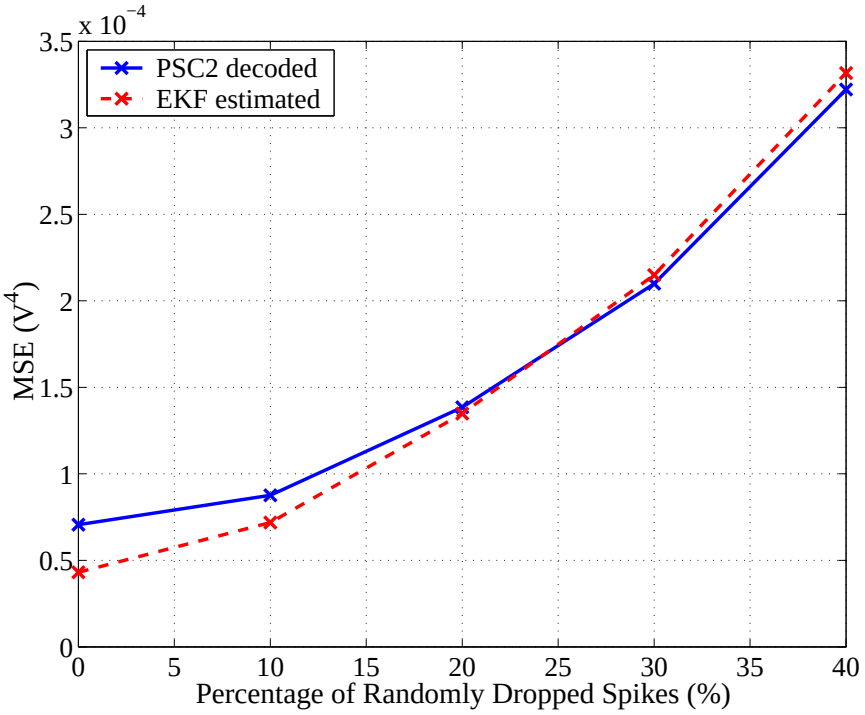

Figure 7.8: System performance in the presence of spike-drop noise for PSC2 decoded and EKF estimated signals.

Once again the MSE increases as a function of the number of spikes dropped and the EKF outperforms the PSC2 filter but only up to the point where 25 % or less of the spikes are dropped randomly. Even though the measurement noise model was not designed to contend explicitly with spikes being dropped it remains a good model for a noisy channel where no more than 25 % of the spikes are lost. However, the model is not as suitable to represent the dropping of spikes incurred as a result of channel noise, as spike-time jitter noise and additional modelling is required when such a channel is indeed present.

7.7 RELATIONSHIP TO PSYCHOACOUSTIC DATA

The auditory system model derived here was based on signal processing steps found in the auditory system of a human and particular attention was given to neural channel and noise model as far as biological plausibility was concerned. In order to compare the auditory system model to a real auditory system, experiments have to be made and measured psychoacoustic data compared to simulated results. One such an experiment is to compare the just-noticeable difference (JND) in intensity discrimination in the auditory system to that of the model designed and simulated here.

The psychoacoustic parameter for intensity discrimination is the JND, which is approximately 1 dB (Zwicker and Fastl, 1990) above the auditory threshold of 20 micro-Pascal at 300 Hz. As was shown by Siebert (1970), the modelled standard deviation in the intensity of the input signal can be equated to the psychoacoustic measured JND in the intensity discrimination.

In this case an input signal of the form

$$x(t) = a \cdot sin(2\pi f t), \tag{7.17}$$

is chosen where a is the intensity of the input signal and f the frequency, where f is taken to be at 300 Hz, as up to now the bandpass filter was centred around 300 Hz. A relationship has to be found that relates the estimated power signal to the intensity of the signal $x(t)$ and to compare this to the variance in the estimated acoustic intensity. The only way to associate the SPL to the auditory system model is the common denominator, the spike rate.

In the case of a typical neuron (Colburn, 1973) the intensity can be related to the SPL by

$$SPL\ (dB) = 10 \cdot log_{10}\left(\frac{a}{\text{Auditory threshold}}\right), \tag{7.18}$$

where a is the intensity (measured in watts/cm^2) and the auditory threshold is $0.97e \cdot 10^{-12}$ watts/cm^2 or 20 micro-Pascal. The intensity a, measured in volts, can be related to the

average spike rate $\bar{r}$ by

$$\bar{r} = MaxRate \cdot \frac{a}{\sqrt{2}\ (0.5 \cdot a^2 + 0.15 \cdot 10^{-15})^{1/2}}. \tag{7.19}$$

Similarly to *Equation (7.19)*, in the case of the LIF neuron (see *Subsection 5.4.1*), the intensity of the signal $x(t)$ can be related to the spike rate by first squaring and then low-pass filtering it such that

$$\begin{aligned} z(t) &= 10 \cdot \text{LPF}\left[x^2(t)\right] \\ &= 10 \cdot a^2/R_{norm}. \end{aligned} \tag{7.20}$$

However, since the input to the LIF neuron in *Equation (5.19)* is limited to between 0 and 57.2 μA, *Equation (7.20)* has to be normalised by the factor $10/R_{norm}$. The normalised intensity $z(t)$ can now be related to the spike rate in the LIF neuron by

$$r(t) = \frac{1}{-\tau_{RC} \cdot log_e\left[1 - \dfrac{V_{th}}{R(I_{bias} + \mathbf{g}z(t))}\right]}. \tag{7.21}$$

The intensity/spike-rate relationship of a typical neuron, given by *Equation (7.19)*, can be compared to that of the LIF neuron, given by *Equation (7.21)*, where the units for the intensity of the LIF neuron are unspecified.

The results are shown in *Figure 7.9*.

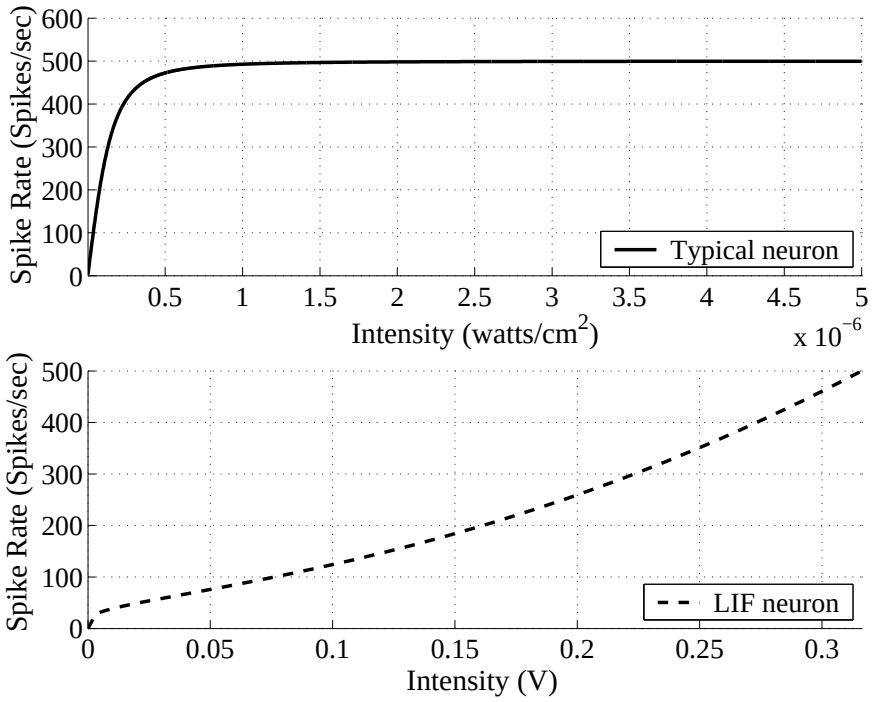

Figure 7.9: Spike rate vs. input intensity comparison between typical neuron and LIF neuron.

It is evident from *Figure 7.9* that in the case of a typical neuron as far as intensities are concerned, the spike rate increases rapidly at increasing, but low, intensities where its most sensitive region is. The LIF neuron on the other hand responds in a linear manner to an increase in intensity.

A 40 dB SPL is chosen as the point of operation and in the case of a typical neuron (from *Equation (7.18)*) results in a constant spike rate of 27.93 spikes per second. This spike rate corresponds to an intensity a of $2.90 \cdot 10^{-3}$ V in the LIF neuron (from *Figure 7.9* bottom). The sinusoid of *Equation (7.17)* is applied to the auditory system model in *Figure 7.2*, where $x(t)$ represents the message $m(t)$ in *Figure 7.2*a. The sinusoidal input results in a constant power signal over time, which is preceded by a transient response phase. In order to correctly calculate the variance in the estimate, the transient response needs to be disregarded and enough time allowed for settling of the estimated power. In this case the setting time was no longer than 0.5 s and the power signal was taken after 0.5 s to determine the variance in the power signal. The variance in the estimated power was found to be $2.29 \cdot 10^{-12}$ V^2, which translates into a standard deviation in the input intensity of $1.23 \cdot 10^{-3}$ V. To translate the deviation in input intensity of the LIF neuron to that of a typical neuron, the deviation in input intensity needs to be related to a deviation in spike rate. The input

intensity of the LIF neuron with the additional standard deviation has now changed from $2.90 \cdot 10^{-3}$ V to $1.67 \cdot 10^{-3}$ V, which translates into a spike rate of 30.33 spikes per second instead of 27.93 spikes per second. This spike rate corresponds to an intensity of $10.53e \cdot 10^{-9}$ Pascal or 40.36 dB SPL. The standard deviation, under noiseless conditions, is thus 0.36 dB, which corresponds very well with the measured psychoacoustic intensity JND of around 1 dB.

Similarly the modelled standard deviation in the intensity of the input signal can be found for a range of SPLs as was done by Zwicker and Fastl (1990). It was shown that the JND or just-noticeable difference in level (JNDL) varies with a varying SPL and can lie anywhere between 2 dB (10 dB SPL) to 0.2 dB (100 dB SPL) for a 1 kHz tone.

In the simulated auditory system of *Figure 7.1*a the dependence of the JND in intensity discrimination on the SPL is shown in *Figure 7.10* over a range of 35 dB to 65 dB SPL.

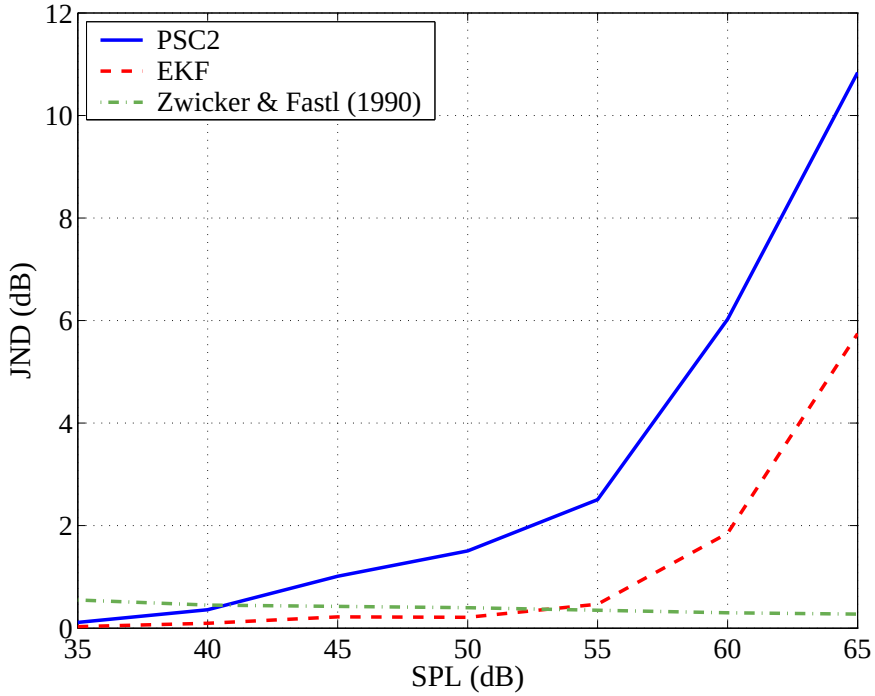

Figure 7.10: The JND in intensity discrimination as a function of the SPL for the simulated auditory signal model for one of the l channels centred around 300 Hz (Compared to psychoacoustic data from Zwicker & Fastl (1990)).

Figure 7.10 indicates that the intensity discrimination in the simulated auditory system model after the neural channel (noiseless) and PSC2 decoding *increases* rather than decreases with SPL as compared to the psychoacoustic data in Zwicker and Fastl (1990). This can be at-

tributed to the particular MSE optimisation technique used in *Section 5.4* to determine the best PSC2 decoding filter. The PSC2 decoding filter is optimised for both the transient and the steady-state response where its optimal bandwidth lies between 5 Hz and 10 Hz. Power signals with a bandwidth below and above this range, such as in this case, result in inferior decoding performance on part of the PSC2 decoding filter since the filter cannot track fast and slow changes in the estimated signal equally well. Thus the standard deviation in intensity discrimination in the auditory system model found was is in fact the *worst case* scenario. Nevertheless, this experiment indicated how close the auditory system model designed and simulated in this dissertation comes to a real auditory system as far as intensity discrimination is concerned.

Another interesting observation that can be made from *Figure 7.10* is the fact that if an EKF is used to estimate the power in the auditory system model from *Figure 7.1*a, the JND is always smaller than that of the PSC2 filtered signal on its own. Although the *trend* of the model output (after PSC2 decoding filter and again after EKF estimator) does not follow the data from Zwicker and Fastl (1990), the *magnitude* is predicted accurately over a wide range of intensities. At 40 dB SPL the JND in intensity discrimination of the PSC2 decoding filter best matches that of the data while at 53.1 dB SPL the EKF best matches the data. From the results in *Figure 7.10* it is also evident that the EKF estimator follows the data over a much wider range of input intensities (35 dB to 53.1 dB) than the PSC2 decoding filter on its own (35 dB to 40.6 dB). These findings are significant in that they underline the necessity of having an *additional* estimator at the output of the PSC2 decoding filter.

A final distinguishing factor between the JND in intensity discrimination found here and that of Zwicker and Fastl (1990) is the fact that neural channel noise was not taken into account in the first case. It can be assumed that some noise existed on the neural channels in the human subject during the gathering of the psychoacoustic data. However, no channel noise was added to the JND in intensity discrimination experiments in this dissertation. The decreasing trend of the JND in intensity discrimination as a function of the SPL found by Zwicker and Fastl (1990) can most probably ascribed to an increase in the SNR in the auditory system.

7.8 DISCUSSION

In this chapter an EKF was initialised and applied to the estimation problem determining the power present in a particular spectral band before neural transmission over a noisy channel takes place. The result was that the EKF was able to estimate the power better than the PSC2 decoder on its own, under all conditions of noise, even though the EKF measurement noise model was optimised for only a very particular channel noise model and noise variance. The EKF only required a noisy measurement after PSC2 decoding to achieve this. Over a 4 s interval (16 000 consecutive estimations) the EKF remained stable and converged to a minimum mean-squared error solution. Each functional unit was both completely observable and controllable. The concatenation of all functional units within the EKF however resulted in an ill-conditioned state transition matrix since all state variables are projected through the last state variable, in this case the power. It is advised to keep the order as low as possible to avoid an ill-conditioned problem. This way the condition number is kept low resulting in a better conditioned problem and hence more predictable system.

Modelling the spike-jitter variance and spike loss as a white Gaussian noise source in the EKF was indeed a suitable, though suboptimal one. This is evident by the large variance ($R_k = 3.5$) that was required for estimation purposes. A coloured measurement noise model however would most likely increase the performance of the filter. Nevertheless, the EKF still performed better in the presence of noise than just using a PSC2 decoder on its own.

When the auditory system model was compared to a real auditory system it was found that the JND in intensity discrimination of the simulated auditory system model found by PSC2 filtering alone performed worse than when the EKF estimated the power signal. In fact for a SPL above 40.6 dB (up to 53.1 dB) the PSC2 filter could not explain the small JND in intensity discrimination found by Zwicker and Fastl (1990) but the auditory system model with EKF estimation could. This result suggest that estimation of signals in the auditory system is vital for accurate auditory perception, requiring advanced estimation techniques such as the Kalman filter.

CHAPTER 8

CONCLUSION

8.1 INTRODUCTION

To understand the neural processing that takes place in the CANS and possibly the CNS, is no trivial feat since the signal processing steps have only been partly and selectively uncovered in the field of neuroscience. In particular, reliable and swift reactions on the part of the CNS when faced with changing input stimuli pose an interesting question as far as modelling of such a process is concerned.

As was mentioned in *Section 1.1* and *Section 1.2*, a concept called the *analysis-by-synthesis method* was successfully applied previously to a motor control problem. The *analysis-by-synthesis method* assumes that an internal model, emulating the behaviour of a particular system, assists with the estimation and decision making process, which could explain the reliability and quick reaction times such as in the case of the motor control problem investigated by Miall et al. (1993) and Wolpert et al. (1995). In this dissertation the same principles were applied to a completely different neural system, the CANS. In particular the primary research question that was asked was: *Does the presence of internal models as was discussed suggest that it is a universal neurophysiologic method for correctly estimating a state in a quick and reliable way? If so what are the biological implications thereof?*.

In this dissertation the question was addressed by

1. First designing a complete model of the auditory system up to and including the inner haircells in *Chapter 4* and *Chapter 5*.

2. Then designing an appropriate EKF in *Chapter 6*. The estimator contained a biologically plausible internal model and was used to estimate the power contained within a particular band, given a noisy transmission medium (neural channel).

3. The estimation performance gained by using an additional estimator (EKF) was investigated relative to only using PSC filters (see *Chapter 7*).

4. Finally, in *Section 7.7* the auditory system model with matching EKF was compared to the biology via psychoacoustic data.

8.2 DISCUSSION OF HYPOTHESES AND RESEARCH QUESTION

In *Chapter 1* the hypotheses made were that

1. there exists an internal model in the auditory system, that by means of an analysis-by-synthesis implementation, can explain the functioning of the system and

2. it is possible to take an analysis-by-synthesis design approach as compared to the more traditional analysis approach and by doing so gain a better understanding of the auditory system and neural coding.

In order to test the hypotheses the second proposition had to be assumed true before the first proposition could be tested. In this dissertation it was shown that an EKF containing a forward-predictive model of the auditory system and a spike coder/decoder model as part of its internal model could indeed explain some characteristics of the auditory system when compared to the JND in intensity discrimination (*Section 7.7*). It was apparent from the results obtained that while a biologically plausible decoding filter such as the PSC2 filter could only partially explain such low JNDs in intensity discrimination (from 0 dB up to 40.6 dB), the EKF could do so over a wider range of SPLs (from 0 dB up to 53.1 dB). It was also shown that for different levels of channel noise (spike-time jitter noise and loss of spikes up to 25 %) the EKF performed better in the MSE sense than the PSC2 decoder.

Combining these findings with the research done by Miall et al. (1993) and Wolpert et al. (1995) where it was shown that an analysis-by-synthesis implementation using a Kalman filter (with forward-predictive model) can be used to describe the coordinated movement of limbs by means of the cerebellum in humans (see *Figure 1.2*), it appears that an analysis-by-synthesis implementation by means of a forward-predictive model forming part of the internal model of the Kalman filter could be a universal neurophysiological method for the correct estimation of a desired state.

8.3 RESEARCH CONTRIBUTION

The contributions that were made are summarised below:

1. A biologically plausible auditory system model representing the auditory system from the outer ear right through to the haircells coding the basilar membrane movement was designed using functional units similar to the ones found in multichannel cochlear implant designs (Meyer-Bäse and Scheich, 1997; Loizou, 1999) and other auditory system models (Flanagan, 1972).

2. Tracking of discrete events or point processes (Snyder, 1975; Gray et al., 1994) as is required for the coder/decoder functional unit in *Figure 7.2*, or formants in speech (Niranjan and Cox, 1994) as required for the source functional unit in *Figure 7.2* by means of Kalman filters is nothing new. However, using similar principles to track a source signal (formant or other) as it travels through the *complete* auditory system up to and including the mechanical to neural transduction and the estimation of a power spectral map as discussed in *Section 4.5*, is unique as far as is known.

3. An analysis-by-synthesis implementation using a Kalman filter with a forward-predictive model was previously applied to a problem concerning the coordinated movement of limbs by means of the cerebellum (Miall et al., 1993; Wolpert et al., 1995). It was argued that the CNS internally simulates the behaviour of the motor system in order to quickly and reliably estimate a state without having to rely solely on slow and unreliable sensory feedback. The same concept was, as far as is known, applied for the first time to the auditory system here.

4. A problem not adequately addressed in the literature (Rieke et al., 1997; Eliasmith and Anderson, 2003) is the correct scaling of PSC1 and PSC2 decoding filters. Correct

scaling in the biology is performed via trained weighting factors in the dendritic trees. For simulation purposes the PSC1 and PSC2 time constants were assumed to be fixed by the biology to 4 ms. Thus only the gain of the PSC filters remained as a tuning parameter. For a rough estimate, the DC gains of the PSC filters were matched to that of the DC gain of the optimal filter, which was derived beforehand. The gain of the PSC filters was then tuned so that the PSC decoded signal (noiseless channel) was the MMSE estimate as compared to the true signal. The MMSE gain obtained was then used as the default weight for one particular channel, for all experiments.

5. As far as is known, the stochastic model (periodic random process) represents the voiced sounds in *Section 4.3* more accurately than the ones found in the literature (Moore, 1972; Bar-Shalom and Li, 1993; Haykin, 1994; Niranjan and Cox, 1994). In addition, the stochastic source signal can model the envelope of a formant frequency with a non-zero mean allowing for more precise modelling and simulation and ultimately better MMSE estimates.

6. The estimator model derived in *Chapter 6* is based on the exact model of the complete auditory system model designed in *Chapter 4* and *Chapter 5*. Since no simplifications were made to the stochastic model representing the auditory system model before implementing it in the internal model of the EKF, the estimator is the best MMSE estimator for this particular model. Since the estimator estimates the lowest hierarchial level information, the power contained within a particular spectral band, this information can be used by other higher-level estimators. It thus provides the critical foundation found at *Levels 1&2* in *Figure 8.1* from which all higher hierarchial level systems can make use.

8.4 IMPLICATIONS

8.4.1 Implications for the Biology

Given the findings in *Chapter 7*, the implications for the biology are as follows (Note that the biological implementation of such an estimator at neural level was of no interest but its implementation at system level was.):

The results suggest that a Kalman filter with forward-predictive model of the auditory system model (see *Figure 1.5*) increases the reliability of the estimated power in almost all cases

(except when more than 25 % of the spikes are dropped). Since the dropping of spikes was not modelled explicitly in the EKF the poor performance under such conditions are of little concern at present. Nevertheless, the increased reliability offered by an additional EKF underlines its plausibility as a tool used by the auditory system to better estimate signals in the presence of noise. In fact in *Section 7.7* the necessity of an EKF was identified since the PSC2 filter on its own could not explain the small measured psychoacoustic values of the JND in intensity discrimination found by Zwicker and Fastl (1990) but the additional EKF could. It thus seems that an additional estimator is much more of a necessity than just being an additional tool available to the auditory system. The choice of using specifically an EKF from the family of estimators that use an internal model of the signal origin is more a matter of preference although the ease with which it can be implemented by a digital simulation tool such as MATLAB and its robustness and versatility makes it a better option than a Wiener filter for example.

An advantage of having taken an analysis-by-synthesis design approach (i.e. modelling a concept rather than mimicking it functionality) was that neural coding issues were avoided such as what neural code should be used to code the power signal *etc.* Choosing a LIF coder for spike coding and a linear PSC decoder for signal extraction, demonstrated that it is possible to evaluate neural systems by making assumptions about the neural channel based on the limited knowledge of neural codes available today. Nevertheless, it did not impede the design of the higher hierarchial level systems such as the Kalman filter estimator but merely placed restrictions on the transmission capabilities of the complete system. By doing so interesting issues were raised such as how well can the decoder and EKF can estimate the original power signal given a particular neural channel corrupted by noise. This suggests that the current trend to find the "correct" neural code for a particular neural system in the CANS or CNS in general should be superseded by first identifying the *need* of the system as a whole and what information is found at various points along the signal processing path. Only then can deductions be made of what information is hidden in the spike trains found in that particular neural system.

Another implication of the findings in *Chapter 7* for the biology is that Kalman based estimators, such as the one described, only explain lower hierarchical level functioning of the CANS as discussed in *Section 1.6*.

The Kalman filter, based on a process and measurement model, estimates the power for a particular spectral band within the entire spectrum (*Level 1&2* in *Figure 8.1*).

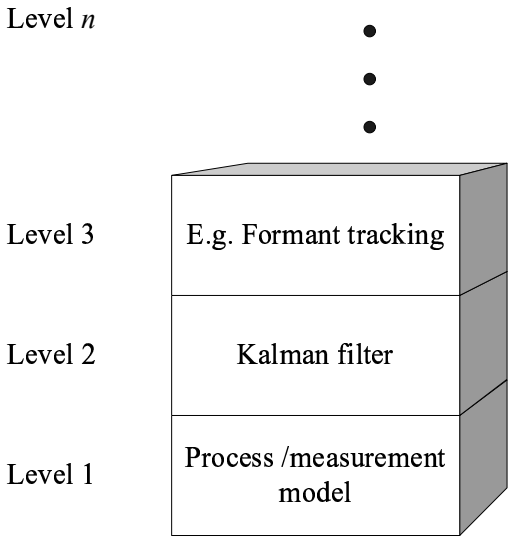

Figure 8.1: Kalman filtering applications for different hierarchial level problems.

Once the power is reliably estimated a spectral map can be generated. However, for the Kalman filter to operate satisfactory the assumption was made that knowledge of the process noise and measurement noise is available, which is most certainly not the case in the biology, due to continuous changes in the external stimuli and the characteristics of the channel.

The Kalman filter in the biology thus requires another lower hierarchial level unit to obtain the stimuli and channel characteristics and possibly a higher hierarchial level observer unit to relay the information as shown in *Figure 8.2*.

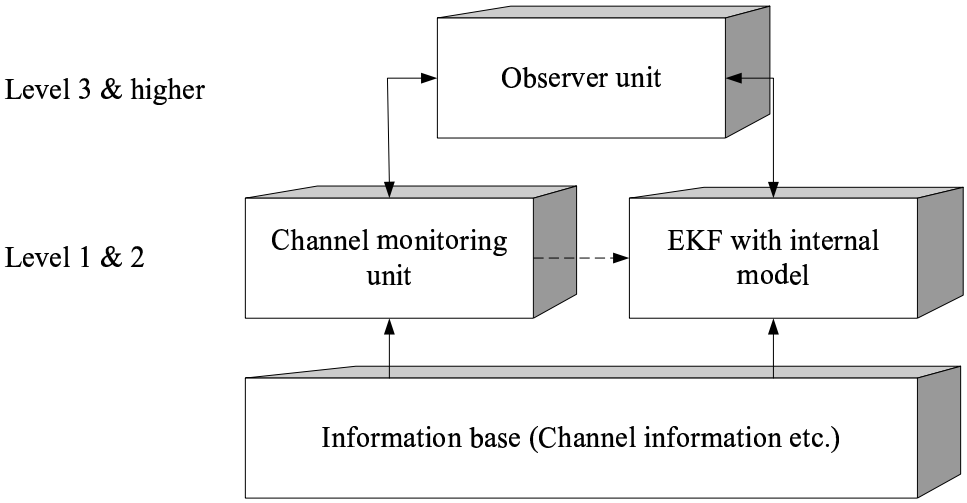

Figure 8.2: In order for the EKF to operate effectively, proper initialisation by an observer unit is required. For this the observer unit however requires additional information collected by the channel monitoring system.

Nevertheless, the lower hierarchial level model provides the foundation on which all higher hierarchial level observer and estimator systems can be based on in the future, like for example multiple-formant tracking (Niranjan and Cox, 1994).

The strong and weak points of the models are discussed next.

8.4.2 Strong and Weak Points of the Models

The strong points of the models are:

1. Since the auditory system model was modelled as accurately as possible given the biological restrictions placed on it in *Chapter 1*, a very accurate stochastic equivalent model could be derived that formed part of the internal model of the EKF. Since the internal model is such a close representation of the auditory system model very little injection or fictitious process noise was required for successful estimation (see *Subsection 6.7.5*). This in turn suggests that a very small difference exists between the auditory

system model and the internal model of the EKF, which allows for a finer estimation resolution or smaller estimator bandwidth.

2. The design of the forward-predictive model was kept modular by representing each functional unit by a stochastic model. This allows for the removal of any of the present functional units or for appending additional ones in order to realize alternative auditory system models.

3. The periodic random process utilised as the stochastic model representing the voiced sounds has a wide scope of applications. Not only is the model suitable for stochastically representing voiced sounds (a group of formant frequencies) but it can also be used to simulate any single formant if required. A further advantage of the periodic random process is its ability to approximate the non-stationary behaviour of the source signal model in *Section 4.3*, where the source signal remains constant over a period of 20 ms before assuming a new random amplitude, frequency and phase.

4. The auditory system model and estimator model completely represent the first two hierarchial levels of *Figure 8.1* making it suitable to be used as platform for higher hierarchial level systems such as a Kalman filter tracking only formant frequencies as was done by Niranjan and Cox (1994) and Lu and Doerschuk (1996).

5. The EKF copes well with neural channel noise for which it was not necessarily optimised for. In *Subsection 7.6.2* it was evident from the results obtained that the EKF performance, as far as the MSE is concerned, degrades only slightly compared to the PSC2 filter over a wide range of spike-time jitter noise standard deviations. In fact for a non-Gaussian noise source such as the dropping of spikes, the EKF performed better than the PSC2 decoding filter in the case where less than 25 % of the spikes were dropped. The EKF however never became unstable and diverged in the presence of either noise source (spike-time jitter and loss of spikes) at any stage.

The potentially weak points of the models are:

1. For a single neural channel the number of state variables that need to be kept track of for the simulation of the forward-predictive model is large (14 state variables, see also *Table (6.5)*). With such a large filter-order the conditioning of the observability and controllability matrices became a relevant point of concern as was shown in *Section 7.6*. The large filter-order of the EKF also increases the complexity of the simulations

and thus the simulation times. This places a restriction on the amount of channels that can be simulated simultaneously and the simulation of a large number of channels is thus not advisable.

2. No specific shaping filter was chosen to represent the two noise processes found on the neural channel (spike-time jitter and loss of spikes). The assumption that these two noise sources could be modelled by a simple Gaussian noise process was a crude one, requiring a large measurement noise variance for proper operation thus limiting the bandwidth of the estimator (i.e. limiting the accuracy of the EKF in estimating the true signal).

8.5 FUTURE RESEARCH POSSIBILITIES

Deriving the stochastic equivalent model from the auditory system model (see *Chapter 6*) was a time-consuming task as an equivalent stochastic model had to be found for each functional unit of the auditory system model, which however, resulted in a very accurate internal model in the EKF (see results in *Section 7.6*). In future it would be interesting to investigate how a simpler, lower-order EKF compares to the EKF designed in this dissertation as far as MSE performance is concerned. The lower-order system and measurement models could be found by using more traditional power estimation techniques such as the ones mentioned in *Addendum A.1.*

In this dissertation the number of neural channels were limited to a single neural channel due to the overall computational complexity of the stochastic equivalent model in the EKF. However, since there are many neural channels transmitting information coded by the hair-cells and overlaps in spectra exists, as depicted in *Figure 4.3* and discussed in *Section 3.5* and *Section 3.6*, an investigation as to how such spectral overlaps can be exploited by decoders and estimators, can be made. Since cross-correlations exist between adjacent neural channels (i.e. parts of the information are shared between them) it allows for estimation of the power spectral map using *sensor fusion principles* (Klein, 1993; Arras, Tomatis, Jensen and Siegwart, 2001) and Kalman filtering techniques. Once a multi-channel implementation is in place that allows for coding and transmission of all power components in the spectral map, it would be interesting to see how well this spectral map approximates the true source signal. This can be assessed by converting the spectral map, which consists of discrete powers found at frequencies determined by the location of the centre-frequencies of each bandpass filter in

the filter bank, to a temporal signal by means of Fourier series expansions and comparing it to the true source signal.

Finally, in *Section 6.7*, channel noise and decoding errors were grouped as part of the measurement noise model. It was assumed for the sake of simplicity, that the two channel-noise processes can be modelled by a white measurement noise process with zero mean and some variance. It was however evident from the large measurement variance required for the successful operation of the EKF, that this was a rather crude assumption. Even though the EKF outperformed the PSC2 decoding filter in the presence of channel noise and decoding errors in almost all instances, such a large measurement noise variance artificially keeps the estimator bandwidth wider than necessary. A better model for the different noise sources would ultimately allow for a better minimum EKF bandwidth, resulting in a better estimate.

REFERENCES

Abbott, L. F. (1994). Decoding neuronal firing and modeling neural networks, *Quarterly Review of Biophysics,* **27**(27): 291–331.

Allen, J. B. (1985). Cochlear modeling, *IEEE ASSP,* pp. 3–29.

Alstrøm, P., Beierholm, U., Nielsen, C. D., Ryge, J. and Kiehn, O. (2002). Reliability of neural encoding, *Physica A,* **314**: 61–68.

Anderson, B. D. O. and Moore, J. B. (1979). *Optimal Filtering*, Prentice-Hall Inc., Englewood Cliffs, New Jersey.

Arras, K. O., Tomatis, N., Jensen, B. T. and Siegwart, R. (2001). Multisensor on-the-fly localization: Precision and reliability for applications, *Robotics and Autonomous Systems,* **34**: 131–143.

Azouz, R. and Gray, C. M. (1999). Cellular mechanisms contributing to response variability of cortical neurons in vivo, *Neuroscience,* **19**: 2209–2223.

Azouz, R. and Gray, C. M. (2000). Dynamic spike threshold reveals a mechanism for synaptic coincidence detection in cortical neurons in vivo, *Proceedings of the National Academy of Science of the USA,* **97**: 8110–8115.

Bar-Shalom, Y. and Li, X.-R. (1993). *Estimation and Tracking: Principles, Techniques and Software*, Artech House.

Bialek, W., Rieke, F., de Ruyter van Steveninck, R. R. and Warland, D. (1991). Reading a neural code, *Science,* **252**: 1854–1857.

Blamey, P., Arndt, P., Bergeron, F., Bredberg, G., Brimacombe, J. and et al. (1996). Factors affecting auditory performance of postlinguistically deaf adults using cochlear implants, *Audiology and Neuro-Otology,* **1**: 293–306.

Bozic, S. M. (1979). *Digital and Kalman Filtering*, Edward Arnold, Hodder and Stoughton, London.

Brenner, N., Bialek, W. and de Ruyter van Steveninck, R. R. (2000). Adaptive rescaling maximizes information transmission, *Neuron,* **26**: 695–702.

Brown, R. G. and Hwang, P. Y. C. (1997). *Introduction To Random Signals And Applied Kalman Filtering*, John Wiley and Sons, Inc., New York.

Brunel, N. and Nadal, J. (1998). Mutual information, Fisher information, and population coding, *IEEE Neural Computation,* **10**: 1731–1757.

Burghes, D. and Graham, A. (1980). *Introduction to Control Theory including Optimal Control*, Ellis Horwood Ltd., West Sussex, England.

Cecchi, G., Sigman, M., Alonso, J.-M., Martinez, L., Chialvo, D. R. and Magnasco, M. O. (2000). Noise in neurons is message dependent, *Proceedings of the National Academy of Science of the USA,* **97**(10): 5557–5561.

Chen, G., Wang, J. and Shieh, L. S. (1997). Interval Kalman filtering, *IEEE Transactions on Aerospace and Electronic Systems,* **33**(1): 250–259.

Chen, G., Xie, Q. and Shieh, L. S. (1998). Fuzzy Kalman filtering, *Journal of Information Sciences,* **109**: 197–209.

Chui, C. K. and Chen, G. (1987). *Kalman Filtering with Real-Time Applications*, 2nd edn, Springer Verlag, Berlin.

Colburn, H. S. (1973). Theory of binaural interaction based on auditory-nerve data. I. General strategy and preliminary results on interaural discrimination, *The Journal of the Acoustical Society of America,* **54**(6): 1458–1470.

Cole, R. A. (ed.) (1980). *Perception and Production of Fluent Speech*, Lawrence Erlbaum Associates, Publishers, New Jersey.

Cooper, H. (ed.) (1991). *Practical aspects of audiology. Cochlear implants, a practical guide*, Whurr Publishers Ltd., London.

Daubechies, I. and Maes, S. (1996). A nonlinear squeezing of the continuous wavelet transform based on auditory nerve models, *Wavelets in Medicine and Biology,* pp. 527–546.

Deco, G. and Schürmann, B. (1999). Spatiotemporal coding in the cortex: Information flow-based learning in spiking neural networks, *IEEE Neural Computation,* **11**(4): 919–934.

DeRusso, P. M., Roy, R. J. and Close, C. M. (1965). *State Variables for Engineers*, John Wiley and Sons, Inc., New York.

Deutsch, S. and Deutsch, A. (1993). *Understanding the Nervous System, An Engineering Approach*, IEEE Press, New Jersey.

Doh-Suk, K., Lee, S. and Rhee, M. (1999). Auditory processing of speech signals for robust speech recognition in real-world noisy environments, *IEEE Transactions on Speech and Audio Processing,* **7**(1): 55–69.

Eggermont, J. J. (2001). Between sound and perception: reviewing the search for a neural code, *Hearing Research,* **157**: 1–42.

Eliasmith, C. and Anderson, C. H. (2003). *Neural Engineering, Computation, Representation, and Dynamics in Neurobiological Systems*, The MIT Press, Cambridge, MA.

Fairhall, A. L., Lewen, G. D., Bialek, W. and de Ruyter van Steveninck, R. R. (2001). Efficiency and ambiguity in an adaptive neural code, *Nature,* **412**: 787–792.

Flanagan, J. L. (1972). *Speech Analysis, Synthesis and Perception*, Springer-Verlag, Berlin.

Frisina, R. D. (2001). Subcortical neural coding mechanism for auditory temporal processing, *Hearing Research,* **158**: 1–27.

Furuta, K., Sano, A. and Atherton, D. (1988). *State Variable Methods In Automatic Control*, John Wiley and Sons, Inc., New York.

Gelb, A. (ed.) (1984). *Applied Optimal Estimation*, MIT Press, MA, Cambridge.

Gray, D. A., Slocumb, B. J. and Elton, S. D. (1994). Parameter estimation for periodic discrete even processes, *Proceedings ICASSP '94*, Vol. 4, pp. 93–96.

Greenberg, S. (1997). *Diversity of Auditory Mechanics*, World Scientific Publishers, Singapore, chapter The Significance of the Cochlear Traveling Wave for Theories of Frequency Analysis and Pitch.

Grewal, M. S. and Andrews, A. P. (2001). *Kalman Filtering, Theory and Practice Using MATLAB*, John Wiley and Sons, Inc., New York.

Hammarberg, B. (2002). *A Signal Processing Approach to Practical Neurophysiology: A Search for Improved Methods in Clinical Routine and Research*, PhD thesis, Uppsala University, Sweden.

Hanekom, J. J. and Krüger, J. J. (2000). A model of frequency discrimination with optimal processing of the auditory never spike interval, *Hearing Research,* **151**: 188–204.

Hayes, M. H. (1996). *Statistical Digital Signal Processing and Modeling*, John Wiley and Sons, Inc., New York.

Haykin, S. (1994). *Communication Systems*, John Wiley and Sons, Inc., New York.

Hodgkin, A. L. and Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve, *Journal of Physiology,* **117**: 500–544.

Houzel, J., Milleret, C. and Innocenti, G. (1994). Morphology of callosal axons interconnecting areas 17 and 18 of the cat, *European Journal of Neuroscience,* **6**: 898–917.

Humayun, M. S., Weiland, J. D., Fujii, G. Y., Greenberg, R., Williamson, R., Little, J., Mech, B., Cimmarusti, V., Boemel, G. V., Dagnelie, G. and de Juan Jr., E. (2003). Visual perception in a blind subject with a chronic microelectronic retinal prosthesis, *Vision research,* **43**: 2573–2581.

Irino, T. and Unoki, M. (1999). An analysis/synthesis auditory filterbank based on an IIR implementation of the Gammachirp, *The Journal of the Acoustical Society of Japan,* **20**(5): 397–406.

Ito, K., Suzuki, Y., Toma, M., Shiroma, M. and Kaga, K. (2002). Postlingual collapse of language and its recovery after cochlear implantation, *International Journal of Pediatric Otorhinolaryngology,* **62**: 261–265.

Johnson, D. J. (1996). Point process models of single-neuron discharges, *Journal of Computational Neuroscience,* **3**: 275–299.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems, *Trans. ASME,* pp. 35–45.

Kay, S. M. (1988). *Modern Spectral Estimation*, Prentice-Hall Inc., Englewood Cliffs, New Jersey.

Klein, L. A. (1993). *Sensor and data fusion concepts and applications*, SPIE Optical Engineering Press, Bellingham, Washington, USA.

Kohn, A. F. (1997). Computer simulation of noise resulting from random synaptic activities, *Computational Biological Medicine,* **27**(4): 293–308.

Kolk, W. R. and Lerman, R. A. (1992). *Nonlinear System Dynamics*, Van Nostrand Reinhold, New York.

Kumaresan, R. and Wang, Y. (2001). On representing signals using only timing information, *The Journal of the Acoustical Society of America,* pp. 2421–2439.

Lampl, I., Reichova, I. and Fenster, D. (1999). Synchronous membrane potential fluctuations in neurons of the visual cortex, *Neuron,* **22**: 361–374.

Le-Mgoc, T. and Vo, M. T. (1989). Implementation and performance of the fast Hartley transform, *IEEE Micro,* **9**(5): 20–27.

Lestienne, R. (2001). Spike timing, synchronization and information processing on the sensory side of the central nervous system, *Proceedings in Neurobiology,* pp. 546–591.

Lewis, F. L. (1986). *Optimal Estimation: With An Introduction To Stochastic Control Theory*, John Wiley and Sons, Inc., New York.

Liberman, A. M. and Mattingly, I. G. (1985). The motor theory of speech perception revised, *Cognition,* **21**: 1–36.

Liberman, M. C. (1982). Single-neuron labeling in the cat auditory nerve, *Science* **216**(4551): 1239–1241.

Liberman, M. C. (1984). Single-neuron labeling and chronic cochlear pathology I. Threshold shift and characteristic-frequency shift, *Hearing Research* **16**(1): 33–41.

Liberman, M. C. and Dodds, L. W. (1984a). Single-neuron labeling and chronic cochlear pathology II. Stereocilia damage and alterations of spontaneous discharge rates, *Hearing Research* **16**(1): 43–53.

Liberman, M. C. and Dodds, L. W. (1984b). Single-neuron labeling and chronic cochlear pathology III. Stereocilia damage and alterations of threshold tuning curves, *Hearing Research* **16**(1): 55–74.

Liberman, M. C. and Kiang, N. Y.-S. (1984). Single-neuron labeling and chronic cochlear pathology IV. Stereocilia damage and alterations in rate- and phase-level functions, *Hearing Research* **16**(1): 75–90.

Loizou, P. C. (1999). Introduction to cochlear implants, *IEEE Engineering in Medicine and Biology,* **18**(1): 32–42.

Lu, S. and Doerschuk, P. C. (1996). Nonlinear modeling and processing of speech based on sums of AM-FM formant models, *IEEE Transactions on Signal Processing,* **44**(4): 773–782.

Maass, W. and Bishop, C. (1999). *Pulsed Neural Networks*, MIT Press, Cambridge, MA.

MacKay, D. M. and McCulloch, W. S. (1952). The limiting information capacity of a neuronal link, *Bulletin for Mathematical Biophysics,* **14**: 127.

Marple, S. L. (1987). *Digital Spectral Analysis with Applications*, Prentice-Hall Inc., Englewood Cliffs, New Jersey.

Martignon, L., Deco, G. and Laskey, K. (2000). Neural coding: Higher-order temporal patterns in the neurostatistics of cell assemblies, *IEEE Neural Computation,* **12**(11): 2621–2654.

Masuda, N. and Aihara, K. (2002). Spatiotemporal spike encoding of continuous external signal, *IEEE Neural Computation,* **14**(7): 1599–1628.

Meyer-Bäse, U. and Scheich, H. (1997). An auditory neuron model for cochlea implants, *Aerosense 97, SPIE Press, Bellingham, Washington, USA,* **3072**: 582–593.

Meyer, J.-U. (2002). Retina implant a bioMEMS challenge A, *Sensors and Actuators,* **97-98**: 1–9.

Miall, R. C., Weir, D. J., Wolpert, D. M. and Stein, J. F. (1993). Is the cerebellum a Smith predictor?, *Journal of Motor Behavior,* **25**(3): 203–216.

Moody, D. B., Stebbins, W. C. and Hawkins, J. R. (1976). *Effects of Noise on Hearing*, Raven, New York, chapter Noise-induced hearing loss in the monkey.

Moore, J. B. (1972). An introduction to optimal communication system theory, *Proceedings of IREE,* pp. 209–215.

Nadol, J. B. (1983). Serial section reconstruction of the neural poles of hair cells in the human organ of Corti. I. Inner hair cells, *Laryngoscope* **93**(6): 780–791.

Naumov, B. (2000). *Philosophy of Nonlinear Control Systems*, CRC Press, Boston.

Niranjan, M. and Cox, I. J. (1994). Recursive tracking of formant in speech signals, *ICASSP-94*, IEEE, pp. 205–208.

Nirenberg, S. and Latham, P. E. (1998). Population coding in the retina, *Current Opinion in Neurobiology,* **8**: 488–493.

O'Neil, W. D., Lin, J. C. and Ma, Y.-C. (1986). Estimation and verification of a stochastic neuron model, *IEEE Transactions in Biomedical Engineering,* **BME-33**: 654–666.

Oppenheim, A. V. and Schafer, R. W. (1989). *Discrete-Time Signal Processing*, Prentice-Hall Inc., Englewood Cliffs, New Jersey.

Panzeri, S. and Schultz, S. R. (2001). A unified approach to the study of temporal correlation, and rate coding, *IEEE Neural Computation,* **13**(6): 1311–1350.

Papoulis, A. and Pillai, S. U. (2002). *Probability, Random Variables and Stochastic Processes*, 4th edn, McGraw-Hill Inc., Boston.

Patterson, R. D., Allerhand, M. and Giguere, C. (1995). Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform, *The Journal of the Acoustical Society of America,* **98**: 1890–1895.

Pitts, W. and McCulloch, W. S. (1947). How we know universals, *Bulletin of Mathematical Biophysics,* **9**: 127–147.

Pouget, A., Deneve, S., Ducom, J. and Latham, P. E. (1999). Narrow vs. wide tuning curves: What's best for a population code?, *IEEE: Neural Computation,* **11**: 85–90.

Pouget, A., Zhang, K., Deneve, S. and Latham, P. E. (1998). Statistically efficient estimation using population codes, *IEEE Neural Computation,* **10**: 373–401.

Proakis, J. G. (1995). *Digital Communications*, 3rd edn, McGraw-Hill Inc., New York.

Rao, K. K. and Yin, P. (1990). *Discrete Cosine Transform: Algorithms, Advantages, Applications*, Academic Press, London.

Rasminsky, M. (1978). *Physiology and Pathobiology of Axons*, Raven Press, New York, chapter Physiology of conduction in demyelinated axons, pp. 361–374.

Rieke, F., Warland, D., de Ruyter van Steveninck, R. and Bialek, W. (1997). *Spikes: Exploring the Neural Code*, MIT Press, MA.

Rose, J. E., Brugge, J. F., Anderson, D. and Hind, J. E. (1967). Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey, *Journal of Neurophysiology,* **30**: 769–793.

Rullen, R. V. and Thorpe, S. J. (2001). Rate coding versus temporal order coding: What the retinal ganglion cells tell the visual cortex, *Neural Computation,* **13**(6): 1255–1284.

Saito, A. and Noguchi, A. (1981). Dynamic analysis for an active nerve pulse of the single unmyelinated fiber, *IEEE Transactions in Biomedical Engineering,* **BME-28**: 812–816.

Schreiber, S., Fellous, J. M., Whimer, D., Tiesinga, P. and Sejnowski, T. J. (2003). A new correlation-based measure of spike timing reliability, *Neurocomputing,* **52-54**: 925–931.

Seifert, E., Oswald, M., Burns, U., Vischer, M., Kompis, M. and Haeusler, R. (2002). Changes of voice and articulation in children with cochlear implants, *International Journal of Pediatric Otorhinolaryngology,* **66**: 115–123.

Shannon, C. E. and Weaver, W. (1949). *The Mathematical Theory of Communication*, University of Illinois Press, Illinois, Urbana.

Siebert, W. M. (1970). Frequency discrimination in the auditory system: Place or periodicity mechanism?, *Proceedings of the IEEE,* **58**(5): 723–730.

Smirnakis, S., Berry, M. J., Warland, D., Bialek, W. and Meister, M. (1997). Adaptation of retinal processing to image contrast and spatial scale, *Nature,* **386**: 69–73.

Snyder, D. L. (1975). *Random Point Processes*, John Wiley and Sons, Inc., New York.

Sorenson, H. W. (ed.) (1985). *Kalman Filtering: Theory and Application*, IEEE Press, New York.

Spoendlin, H. (1974). *Facts and Models in Hearing*, Springer Verlag, New York.

Subtelny, J. D. (ed.) (1980). *Speech Assessment and Speech Improvement for the Hearing Impaired*, 1 edn, The Alexander Graham Bell Association for the Deaf, Inc., Washington, D.C. 20007, USA.

Teich, M. C. and Khannan, S. M. (1985). Pulse number distribution for the neural spike train in the cat's auditory nerve, *The Journal of the Acoustical Society of America,* **46**: 1110–1128.

Tiesinga, P. H. E., Fellous, J.-M. and Sejnowski, T. J. (2002). Spike-time reliability of periodically driven integrate-and-fire neurons, *Neurocomputing,* **44-46**: 195–200.

Tobey, E. (1993). *Cochlear Implants: Audiological Foundations*, Singular Publishing Group, Inc, chapter Speech production, pp. 257–316.

van Loan, C. F. (1978). Computing integrals involving the matrix exponential, *IEEE Transactions on Automatic Control,* **23**(3): 395–404.

Van Trees, H. L. (1968-1971). *Detection, Estimation and Modulation Theory*, I, II and III, John Wiley and Sons, Inc., New York.

Verveen, A. A. and DeFelice, L. J. (1974). Membrane noise, *Progress in Biophysics and Molecular Biology* **28**: 189–265.

Verveen, A. A. and Derksen, H. E. (1968). Fluctuation phenomena in nerve membrane, *Proceedings of the IEEE* **56**(6): 906–916.

Vetterli, M. (1992). *Wavelets and Their Application*, Jones & Bartlett Publishers, Boston, chapter Wavelets and Filter Banks for Discrete-Time Signal Processing, pp. 17–49.

Walker, J. S. (1999). *A Primer on Wavelets and their Scientific Application*, Chapman and Hall, CRC Press, Boca Raton, FL.

Warland, D. K., Reinagel, P. and Meister, M. (1997). Decoding visual information from a population of retinal ganglion cells, *Journal of Neurophysiology,* **78**: 2336–2350.

Warren, R. M. (1999). *Auditory Perception. A New Analysis and Synthesis*, Cambridge University Press, UK.

Watkinson, J. (1995). *Compression in video & audio*, Vol. 1, Focal Press, Linacre House, Jordan Hill, Oxford.

Widrow, B. and Streans, S. D. (1985). *Adaptive Signal Processing*, Prentice-Hall Inc., Englewood Cliffs, New Jersey.

Wilke, S. D. and Eurich, C. W. (2001). Representational accuracy of stochastic neural populations, *IEEE Neural Computation,* **14**: 155–189.

Wolpert, D. M., Ghahramani, Z. and Jordan, M. I. (1995). An internal model for sensorimotor integration, *Science,* **269**: 1880–1882.

Yost, W. A. (2000). *Fundamentals of Hearing, an Introduction*, 4th edn, Academic Press, London.

Zemel, R. S., Dayan, P. and Pouget, A. (1998). Probabilistic interpretation of population codes, *Neural Computation,* **10**(2): 403–430.

Zhang, K., Ginzburg, I., McNaughton, B. L. and Sejnowski, T. J. (1998). Interpreting neuronal population activity by reconstruction: A unified framework with application to the hippocampal place cells, *Journal of Neurophysiology,* **79**: 1017–1044.

Ziemer, R. E., Tranter, W. H. and Fannin, D. R. (1998). *Signal & Systems: Continuous and Discrete*, 4th edn, Prentice-Hall Inc., Englewood Cliffs, New Jersey.

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

ADDENDUM A

DERIVATIONS AND EXTRA INFORMATION

A.1 POWER SPECTRUM ESTIMATION

Kalman filters are basically spectral estimators, estimating the transfer function by analysing the autocorrelation function or power spectral density (PSD), where the PSD is simply the Fourier transform of the autocorrelation function. The Kalman filter however needs some idea of what the transfer function looks like in order to estimate it. It may for example assume that the transfer function is of the form

$$H(z) = \frac{b_1 + b_2 z^{-1} + \ldots + b_{n+1} z^{-n}}{1 + a_2 z^{-1} + \ldots + a_{n+1} z^{-n}}, \tag{A.1}$$

in the z-domain. Each of the coefficients (a_n) and (b_n) can assume a value where the number of poles and zeros determines the order of the system. The ARMA model of *Equation (4.42)* is one example of how a stochastic signal can be modelled. However, to find the coefficients and thus the poles and zeros of the transfer function in *Equation (A.1)*, some form of estimation has to be performed. Spectrum estimation by nonparametric estimation methods such as the *Periodogram, Bartlett's method, Welch's method*, *minimum variance spectrum estimation* method, parametric estimation methods such as the *covariance method, Burg Algorithm, AR, MA* and *ARMA* method, frequency estimation methods like the *Eigendecomposition, Blackman-Tukey frequency estimation* and *minimum variance estimation* method can be found in Hayes (1996). All spectrum estimation methods in one form or another try to approximate the autocorrelation function of the underlying process. By using one of these methods the numerator and denominator coefficients of *Equation (A.1)* can be approximated provided the process is *linear* and *stationary*. Adaptive filters (Widrow and Streans, 1985; Hayes, 1996) on the other hand can estimate *non-stationary processes* just like a Kalman

filter. Adaptive filters and spectrum estimation methods allow for alternative approaches to be taken to implement IIR filters as shaping filters.

A.2 AMPLITUDE DEMODULATION

If a message signal $m(t)$ is modulated by a carrier then

$$s(t) = A_c m(t) \cos(2\pi f_c t), \tag{A.2}$$

where A_c is the carrier amplitude and f_c the carrier frequency.

The demodulation process is just the reverse process, where the modulated signal $s(t)$ is again multiplied by the carrier. If coherent detection is not used then an unknown phase difference ϕ exists and the demodulated signal can be written as

$$\begin{aligned} d(t) &= A_c' s(t) \cos(2\pi f_c t + \phi) \\ &= A_c A_c' m(t) \cos(2\pi f_c t) \ \cos(2\pi f_c t + \phi), \end{aligned} \tag{A.3}$$

where A_c' is the demodulating carrier amplitude.

Using the trigonometric identities

$$\begin{aligned} \cos(\alpha \pm \beta) &= \cos\alpha\cos\beta \mp \sin\alpha\sin\beta \\ \cos\alpha\cos\beta &= \frac{1}{2}\left[\cos(\alpha-\beta) + \cos(\alpha+\beta)\right] \\ \cos\alpha\sin\beta &= \frac{1}{2}\left[\sin(\beta-\alpha) + \sin(\beta+\alpha)\right], \end{aligned} \tag{A.4}$$

the demodulated signal becomes

$$\begin{aligned} d(t) &= m(t)\left[\cos\phi\frac{1}{2}\left[1+\cos(4\pi f_c t)\right]-\sin\phi\frac{1}{2}\left[0+\sin(4\pi f_c t)\right]\right] \\ &= A_c A_c'\frac{m(t)}{2}\left[\cos\phi+\cos(4\pi f_c t+\phi)\right]. \end{aligned} \tag{A.5}$$

A.3 OPTIMAL FILTERING

To find a filter that decodes a particular set of signals, or an *ensemble* of signals, rather than just one specific signal, it is required that the input signal is defined in terms of an input probability distribution function. One way of representing this distribution is to use statistically independent Fourier coefficients that can be represented by the Fourier series

$$x(t;\mathbf{A}) = \sum_{n=-(N-1)/2}^{(N-1)/2} A(\omega_n)\; e^{j\omega_n t}. \tag{A.6}$$

Each value of $A(\omega_n)$ in the vector $\mathbf{A}$ is chosen from a Gaussian distribution. *Equation (A.6)* inherently assumes a stationary ensemble in that the statistics are independent of time.

The response of the neurons, in this case a pair of neurons, to the ensemble of signals can be written as

$$\begin{aligned} R(t;\mathbf{A}) &= \sum_{k}^{M_{on}}\delta(t-t_k^{+}(\mathbf{A}))-\sum_{l}^{M_{off}}\delta(t-t_l^{-}(\mathbf{A})) \\ &= \sum_{i=1}^{2}\sum_{k=1}^{M}\phi_i\delta(t-t_{ik}(\mathbf{A})), \end{aligned} \tag{A.7}$$

where $\phi = 1$ for an 'on' neuron and -1 for an 'off' neuron. In order to find the optimal estimator the error between the true signal and the estimated signal has to be minimised.

The error is given by

$$\begin{aligned} MSE &= E\left[[x(t;(A)) - h(t)\cdot R(t;\mathbf{A})]^2\right]_{t,\mathbf{A}} \\ &= E\left[\left[x(t;\mathbf{A}) - \sum_{i,k} h(t - t_{ik}(\mathbf{A}))\right]^2\right]_{t,\mathbf{A}}. \end{aligned} \tag{A.8}$$

Since it is simpler to evaluate the MMSE of *Equation (A.8)* in the frequency domain, the error for every channel can be written, given a pair of on/off neurons as

$$E(\omega_n) = E\left[\frac{1}{2\pi}|A(\omega_n) - h(\omega_n)R(\omega_n;\mathbf{A})|^2\right]_{\mathbf{A}}. \tag{A.9}$$

Eliasmith and Anderson (2003) solve the optimal filter problem by minimising a windowed form of *Equation (A.9)* which in return results in an optimal filter given by

$$h(\omega_n) = \frac{E\left[A(\omega_n)R^*(\omega_n;\mathbf{A})\right]_{\mathbf{A}}}{E\left[|R(\omega_n;\mathbf{A})|^2\right]_{\mathbf{A}}}. \tag{A.10}$$

The reader is referred to Eliasmith and Anderson (2003) for a detailed insight into the specific process.

A.3.1 Information Transmission of a Linear Estimators for Non-linear Systems

To find the optimal linear estimator h given the response of the neurons $R(x)$ to stimulus x, a general expression for the information transmission can be derived by assuming that the signal is drawn from a Gaussian distribution with $N(0,\sigma_x^2)$. The linear estimate can then be found by

$$\hat{x} = hR(x). \tag{A.11}$$

The average mean-squared error can be found by letting

$$\begin{aligned} MSE &= E\left[(x-\hat{x})^2\right]_x \\ &= E\left[(x-hR(x))^2\right]_x, \end{aligned} \tag{A.12}$$

over all possible signals. By setting *Equation (A.12)*'s derivative to zero and solving for h the MMSE is found. Thus, the optimal linear estimator is given by

$$h = \frac{E\left[x \cdot R(x)\right]_x}{E\left[R^2(x)\right]_x}. \tag{A.13}$$

The signal, or explained variance is the variance of the estimate given by

$$\begin{aligned} E\left[\hat{x}^2\right]_x &= h^2 E\left[R^2(x)\right]_x \\ &= \frac{E\left[x \cdot R(x)\right]_x^2}{E\left[R^2(x)\right]_x}. \end{aligned} \tag{A.14}$$

The signal to noise ratio is the ratio of signal variance to the error variance which in this case is

$$SNR = \frac{E\left[\hat{x}^2\right]_x}{E\left[MSE\right]}, \tag{A.15}$$

since a linear decoder was used to decode a non-linearly encoded signal.

The SNR is thus never infinite. For a Gaussian signal and Gaussian noise, the information transmitted by the Gaussian channel is defined as

$$I = \frac{1}{2}\log_2(1+SNR), \tag{A.16}$$

measured in *bits* and by rewriting *Equation (A.15)* the following is obtained:

$$(1 + SNR) = \frac{E\left[x^2\right]_x}{E\left[x^2\right]_x - \dfrac{E\left[x \cdot R(x)\right]_x^2}{E\left[R^2(x)\right]_x}}. \tag{A.17}$$

The signal can generally be defined by a set of independent Fourier coefficients at discrete points in the frequency domain and written in the general form

$$x = \sum_n A(\omega_n)\Phi_n. \tag{A.18}$$

Equation (A.16) can then be rewritten in a general form including the input signal which results in

$$I(\omega_n) = \frac{1}{2}\log_2\left[\frac{E\left[|A(\omega_n)|^2\right]_{\mathbf{A}}}{E\left[|A(\omega_n)|^2\right]_{\mathbf{A}} - \dfrac{|E\left[\mathbf{A}(\omega_n)R^*(\omega_n;\mathbf{A})\right]_{\mathbf{A}}|^2}{\{|R(\omega_n;\mathbf{A})|^2\}_{\mathbf{A}}}}\right]. \tag{A.19}$$

The information rate is defined as the sum of independent information components across the bandwidth of interest and can be expressed as

$$R_I = \frac{1}{2}\frac{\Delta\omega}{2\pi}\sum_n Info(\omega_n). \tag{A.20}$$

The information rate in this case is measured in *bits per second*.

A.4 LEAKY INTEGRATE-AND-FIRE MODEL

The leaky integrate-and-fire (LIF) neuron can be thought of as a current source driving a capacitor in parallel with a resistor. The voltage across the resistor represents the membrane potential. The first-order linear differential equation can be written as

$$I(t) = \frac{u(t)}{R} + C\frac{dV}{dt}. \tag{A.21}$$

If the time constant $\tau_{RC} = RC$, also known as the membrane time constant, then *Equation*

(A.21) can be written as:

$$\tau_{RC}\frac{dV}{dt} = -V(t) + RI(t). \tag{A.22}$$

If it is required that the solution should be of the form

$$V(t) = F(t)\ e^{-t/\tau_{RC}}, \tag{A.23}$$

Equation (A.23) can be substituted into *Equation (A.22)* resulting in

$$\frac{d}{dt}\left(F(t)\ e^{-t/\tau_{RC}}\right) = -\frac{1}{\tau_{RC}}\left(F(t)\ e^{-t/\tau_{RC}} - I(t)R\right). \tag{A.24}$$

By rearranging *Equation (A.24)* the following is obtained:

$$\frac{dF(t)}{dt} = \frac{I(t)R}{\tau_{RC}}\ e^{t/\tau_{RC}}. \tag{A.25}$$

Integrating both sides, $F(t)$ becomes

$$F(t) = \frac{R}{\tau_{RC}}\int_0^t\ e^{t/\tau_{RC}} I(t')\ dt', \tag{A.26}$$

which results in the general form of a convolution integral

$$V(t) = \frac{R}{\tau_{RC}}\int_0^t\ e^{-(t\,-\,t')/\tau_{RC}} I(t')\ dt'. \tag{A.27}$$

The response of the LIF neuron can thus be found by convoluting an exponential function with the current waveform.

If the current source is constant over the period $[0; t]$ then *Equation (A.27)* simplifies to

$$V(t) = RI_0\left[1 - e^{-(t\,-\,t_0)/\tau_{RC}}\right]. \tag{A.28}$$

Equation (A.28) can be used as an approximation if the waveform $I(t)$ is unknown. *Equation (A.27)* can be used to find a numerically solution if the current waveform is not constant.

A.5 FREQUENCY RESPONSE OF PSC FILTERS

Given the exponential kernel

$$x_{PSC1}(t) = \frac{1}{\tau_{RC}}\, e^{-t/\tau_{RC}}, \tag{A.29}$$

and the fact that the Fourier transform can be obtained by

$$G(f) = \int_{-\infty}^{\infty} g(t)\, e^{-j2\pi ft}\, dt, \tag{A.30}$$

the frequency response of the first PSC filter can be found.

Keeping in mind that

$$\int_{-\infty}^{\infty} t\, e^{-at}\, dt = \frac{1}{a^2}\, e^{-at}\left[-at - 1\right], \tag{A.31}$$

the Fourier transform can now be computed by solving

$$\begin{aligned} G(f) &= \frac{1}{\tau_{RC}} \int_{-\infty}^{\infty} e^{-(\frac{1}{\tau_{RC}} + j2\pi f)t}\, dt \\ &= \frac{1}{\tau_{RC}} \left[-\frac{\tau_{RC}}{1 + j\tau_{RC}2\pi f}\, e^{-(\frac{1}{\tau_{RC}} + j2\pi f)t} \right]_0^{\infty} \\ &= \frac{1}{1 + j\tau_{RC}2\pi f}. \end{aligned} \tag{A.32}$$

The area of the *Equation (A.29)* can be calculated as follows:

$$\begin{aligned}\Delta_{PSC1} &= \frac{1}{\tau_{RC}}\int_{-\infty}^{\infty} e^{-t/\tau_{RC}}\, dt \\ &= \frac{1}{\tau_{RC}}\left[\tau_{RC}\, e^{-t/\tau_{RC}}\right]_0^t \\ &= 1 - e^{-t/\tau_{RC}}. \end{aligned} \tag{A.33}$$

When t is sufficiently large or tends to infinity the area of the exponential PSC tends to one. Similarly the Fourier transform of the alpha-function kernel given by

$$x(t) = \frac{t}{\tau_{RC}^2}\, e^{-t/\tau_{RC}} \tag{A.34}$$

can be found by solving

$$\begin{aligned}G(f) &= \frac{1}{\tau_{RC}^2}\int_{-\infty}^{\infty} t\, e^{-(\frac{1}{\tau_{RC}} + j2\pi f)t}\, dt \\ &= \frac{1}{\tau_{RC}^2}\left[\frac{\tau_{RC}^2}{(1+j2\pi f\tau_{RC})^2}\, e^{-(\frac{1}{\tau_{RC}} + j2\pi f)t}(-\left[\frac{1}{\tau_{RC}} + j2\pi f\right] t - 1)\right]_0^{\infty} \\ &= \frac{1}{(1+j2\pi f\tau_{RC})^2}. \end{aligned} \tag{A.35}$$

The area of the second PSC filter can be obtained by solving

$$\begin{aligned}\Delta_{PSC2} &= \frac{1}{\tau_{RC}^2}\int_0^t t\, e^{-t/\tau_{RC}}\, dt \\ &= \frac{1}{\tau_{RC}^2}\left[\tau_{RC}^2\, e^{-t/\tau_{RC}}(-\frac{t}{\tau_{RC}} - 1)\right]_0^t \\ &= \left[1 - (\frac{t}{\tau_{RC}} + 1)\, e^{-t/\tau_{RC}}\right]. \end{aligned} \tag{A.36}$$

When t is sufficiently large or tends to infinity the area of the alpha-kernel PSC tends to one.

The results obtained are summarised in *Table (A.1)*

Table A.1: Frequency and area characteristics of PSC filters.

PSC type	Fourier transform	Area
Exponential	$\frac{1}{1+j2\pi f\tau_{RC}}$	1
Alpha-function	$\frac{1}{(1+j2\pi f\tau_{RC})^2}$	1

A.6 OBSERVABILITY AND CONTROLLABILITY IN CONTROL SYSTEM THEORY

A deterministic, multi-variable, static plant of the form

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{u}_k \\ y_k &= \mathbf{H}\mathbf{x}_k, \end{aligned} \tag{A.37}$$

can be evaluated in terms of its observability and controllability. A deterministic input $\mathbf{u}_k$ drives a plant $\mathbf{A}$, where the state variables are given by $\mathbf{x}_k$. The observability condition defines the ability to determine the states of a system by observing them via the measurement sensitivity matrix $\mathbf{H}$.

Equation (A.37) is said to be completely observable if the *observability matrix* defined as a (n x np) matrix

$$\Xi \triangleq \left[\mathbf{H}^T, \mathbf{A}^T\mathbf{H}^T, \ldots, (\mathbf{A}^T)^{n-1}\mathbf{H}^T\right], \tag{A.38}$$

has rank n (Furuta et al., 1988; Burghes and Graham, 1980; Gelb, 1984; DeRusso et al., 1965).

A system of the form as in *Equation (A.37)* is completely controllable if the *controllability matrix* defined as a (n x nm) matrix

$$\varphi \triangleq \left[\boldsymbol{B}, \mathbf{A}\boldsymbol{B}, \ldots, \mathbf{A}^{n-1}\mathbf{B}\right], \tag{A.39}$$

has rank n (Furuta et al., 1988; Burghes and Graham, 1980; Gelb, 1984; DeRusso et al., 1965).

As was mentioned earlier, observability and controllability are as important in stochastic models as they are in deterministic ones. For stochastic observability it is required that in the absence of process noise, the measurement noise excites all measured states in the system. Conversely, for stochastic controllability it is required that in the absence of measurement noise, the process noise excites all states in the system (See Gelb (1984) for a formal derivation). For a discrete-time system stochastic observability and controllability is defined in *Subsection A.10.2*.

A.7 NARROW-BAND NOISE

Narrow-band systems are quite common in control and communications theory, the primary idea being to apply a narrow-band filter, with a bandwidth just large enough to pass the signal of interest essentially undistorted while inhibiting excessive noise from passing through. The spectral components of narrow-band noise is situated around a centre frequency f_c and can be specified as

$$n(t) = n_I(t)\cos(2\pi f_c t) - n_Q(t)\sin 2(\pi f_c t), \tag{A.40}$$

where $n_I(t)$ and $n_Q(t)$ are independent Gaussian random processes commonly known as in-phase and quadrature components of the signal $n(t)$. If *Equation (A.40)* is converted from rectangular to polar form by substituting

$$\begin{aligned} n_I &= r\ cos\ \phi \qquad \text{and} \\ n_Q &= r\ sin\ \phi, \end{aligned} \tag{A.41}$$

Equation (A.40) can be rewritten as

$$n(t) = r(t)\cos(2\pi f_c t + \phi(t)). \tag{A.42}$$

It should be obvious from *Equation (A.42)* that the narrow-band process is nothing more than a random process with random amplitude and phase as was found in *Section 4.3*.

The corresponding density functions are

$$f_r(r) = \frac{r}{\sigma^2}\, e^{-r^2/2\sigma^2}, \qquad [r \geq 0] \tag{A.43}$$

and

$$f_\phi(\phi) = \begin{cases} \dfrac{1}{2\pi}, & [0 \leq \phi < 2\pi] \\ 0, & [otherwise] \end{cases}$$

where the former is a *Rayleigh* distribution and the latter a *uniform* distribution.

The autocorrelation function of $n(t)$, when passing white-noise with variance σ^2 through an ideal bandpass filter centred at f_c with bandwidth B, is given by

$$\psi(\tau) = 2\sigma^2 B\ \mathrm{sinc}(B\tau)\cos(2\pi f_c \tau), \tag{A.44}$$

where the spectral autocorrelation functions for the in-phase and quadrature components are given by

$$\psi_{N_I}(\tau) = \psi_{N_Q}(\tau) = 2\sigma^2 B\ \mathrm{sinc}(B\tau). \tag{A.45}$$

A.8 GAUSS-MARKOV PROCESSES

A special class of random processes generated by passing white-noise through linear time invariant filters is the family of Gauss-Markov processes. Defined as (Gelb, 1984) a continuous process, $x(t)$ is *first-order Markov* if for every k and

$$t_1 < t_2 < \ldots < t_k, \tag{A.46}$$

it is true that the probability distribution for the process $x(t_k)$ is dependent only on the value of the directly preceding probability $x(t_k)$ (hence first-order) or

$$F\left[x(t_k)|x(t_{k-1}),\ldots,x(t_1))\right] = F\left[x(t_k)|x(t_{k-1})\right]. \tag{A.47}$$

If the process is continuous it can be given by the linear differential equation

$$\frac{dx}{dt} + \beta_1(t)x(t) = \mathrm{w}(t) \tag{A.48}$$

or in the state-space form as

$$\dot{x}(t) = -\beta_1 x(t) + \mathrm{w}(t), \tag{A.49}$$

where $\mathrm{w}(t)$ is a continuous white process with infinite variance.
If $\mathrm{w}(t)$ were to become a Gaussian random process with finite variance then consequently $x(t)$ would become Gaussian and the process would become a first-order Gauss-Markov process with autocorrelation function

$$\psi_X(\tau) = \sigma^2\ e^{-\beta_1|\tau|} + m_X^2, \tag{A.50}$$

with mean m_X and PSD

$$\Psi_X(\omega) = \frac{2\beta_1\sigma^2}{\omega^2 + {\beta_1}^2}. \tag{A.51}$$

For the continuous second-order Markov process it is true that for every k and

$$t_1 < t_2 < \ldots < t_k, \tag{A.52}$$

it is true that

$$F\left[x(t_k)|x(t_{k-1}),\ldots,x(t_1))\right] = F\left[x(t_k)|x(t_{k-1}),x(t_{k-2})\right]. \tag{A.53}$$

An associated differential equation is

$$\frac{d^2x}{dt} + 2\beta_2(t)\frac{dx}{dt} + \beta_2(t)x(t) = \mathrm{w}(t) \tag{A.54}$$

or in state-space formulation

$$\dot{x}(t) = \begin{bmatrix} 0 & 1 \\ -\beta^2 & -2\beta \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \mathrm{w}(t). \tag{A.55}$$

Again, if $\mathrm{w}(t)$ were to become a Gaussian random process with finite variance then consequently $x(t)$ would become Gaussian and the process would become a second-order Gauss-Markov process with autocorrelation function

$$\psi_X(\tau) = \sigma^2(1 + \beta_2|\tau|)\ e^{-\beta_2|\tau|} + m_X^2, \tag{A.56}$$

with mean m_X and PSD

$$\Psi_X(\omega) = \frac{4{\beta_2}^3\sigma^2}{(\omega^2 + {\beta_2}^2)^2}. \tag{A.57}$$

As a higher-order Gauss-Markov process, the n^{th}-order Gauss-Markov process is given by the autocorrelation function

$$\psi_X(\tau) = \sigma^2\ e^{-\beta_n|\tau|} \sum_{k=0}^{n-1} \frac{\Gamma(n)(2\beta_n|\tau|)^{(n-k-1)}}{(2n-2)!k!\Gamma(n-k)} \tag{A.58}$$

and PSD

$$\Psi_X(\omega) = \frac{(2\beta_n)^{2n-1}\left[\Gamma(n)\right]^2}{(2n-2)!\left(\omega^2 + {\beta^2}_n\right)^n}. \tag{A.59}$$

A.9 LINEARISATION IN THE EXTENDED KALMAN FILTER

A.9.1 Linearisation and the Linearised Kalman Filter

If the trajectory of the state estimates are known up to a certain extent, also known as the *nominal* trajectory, where the measurements are of little importance, a linearised Kalman filter can be derived. Due to *a priori* knowledge of the nominal trajectory, partial derivative matrices can be found beforehand, which allow for off-line computation of the Kalman gains, which can be stored and then utilised during the online process. The Kalman filter however expects to have a perfect internal model with a degree of uncertainty provided by the process and measurement noise, but is known to diverge over time as the nominal trajectory becomes less and less accurate (Gelb, 1984).

While the state estimates are updated through their various non-linear functions, the computation of the error covariance matrix and the Kalman gain require linearised approximations of the state variable and the measurement sensitivity. This can be achieved by two completely different methods. The first method expands the functions $\mathbf{f}(\mathbf{x}(t), t)$ and $\mathbf{h}(\mathbf{x}(t), t)$ around a nominal point (linearised Kalman filter (LKF)) or estimated trajectory (EKF) by means of a Taylor series, while the second expands the probability density function $f(x)$ of the random variable x. The latter method will be described in the next section.

The Taylor expansion of the function $\mathbf{f}(x(t), t)$ for the LKF is given by

$$\mathbf{f}(\mathbf{x}(t), t) = \mathbf{f}(\mathbf{x}^*(t), t) + \left.\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\right|_{\mathbf{x}\,=\,\mathbf{x}^*} (\mathbf{x} - \mathbf{x}^*) + \ldots, \tag{A.60}$$

where $\mathbf{x}^*(t)$ is the nominal trajectory and for the EKF the Taylor expansion is given by

$$\mathbf{f}(\mathbf{x}(t), t) = \mathbf{f}(\hat{\mathbf{x}}(t), t) + \left.\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\right|_{\mathbf{x}\,=\,\hat{\mathbf{x}}} (\mathbf{x} - \hat{\mathbf{x}}) + \ldots, \tag{A.61}$$

where $\hat{\mathbf{x}}(t)$ is the estimated trajectory. In a similar manner $\mathbf{h}(\mathbf{x}, t)$ can be expanded. In this implementation of the EKF only the first partial derivative is used. Higher-order filters also known as *iterated Kalman filter* (IKF) (not to be confused with the *interval Kalman filter* found in Chen, Wang and Shieh (1997) and Chen, Xie and Shieh (1998)) truncate the Taylor

series after the second term or further on. The additional terms account for higher-order effects of the non-linearities and improve performance at the cost of an increased computational burden.

A.9.2 Non-linear Estimation by Statistical Linearisation

If the probability density function of $f(x)$ is known, another expansion method can be used to manage non-linearities, which generally result in a more accurate series expansion than the Taylor series (Gelb, 1984). $f(x)$ is approximated by a series expansion of the form

$$f(x) \cong n_0 + n_1 x + n_2 x^2 + \ldots, \tag{A.62}$$

where the best appropriate coefficients n_k are determined. The most frequently used method for choosing the coefficients is to minimise the mean-squared value

$$E\left[(f(x) - n_0 - n_1 x - \ldots - n_m x^m)^2\right]. \tag{A.63}$$

This method does not require derivatives of f as the Taylor series does and non-linear dynamic systems with a large number of non-linearities are thus treated with ease.

A.10 STOCHASTIC OBSERVABILITY AND CONTROLLABILITY

A.10.1 Stochastic Observability

In the absence of process noise and with perfect *a priori* information the condition for complete observability is given by

$$\alpha_1 \mathbf{I} \leq \sum_{i=k-N}^{k} \mathbf{\Phi}^T(i,k)\mathbf{H}_i^T R_i^{-1}\mathbf{H}_i\mathbf{\Phi}(i,k) \leq \alpha_2 \mathbf{I}, \tag{A.64}$$

for some value of $N > 0$, where $\alpha_1 > 0$ and $\alpha_2 > 0$. The summation term in *Equation (A.64)* is the solution for the error covariance matrix. When the linear system is stationary, the criterion of observability can be expressed algebraically by *Equation (A.38)*. For the

continuous case see Gelb (1984).

A.10.2 Stochastic Controllability

In the absence of measurements and with perfect *a priori information* the condition for complete controllability is given by

$$\beta_1 \mathbf{I} \leq \sum_{i=k-N}^{k-1} \mathbf{\Phi}(k, i+1) Q_i \mathbf{\Phi}^T(k, i+1) \leq \beta_2 \mathbf{I}, \tag{A.65}$$

for some value of $N > 0$, where $\beta_1 > 0$ and $\beta_2 > 0$. When the linear system is stationary and Q is positive definite, the criterion of observability can be expressed algebraically by *Equation (A.39)*. For the continuous case see Gelb (1984).