UNIVERSIDAD AUTÓNOMA DE MADRID

ESCUELA POLITÉCNICA SUPERIOR

PROYECTO FIN DE CARRERA

Speech Enhancement using Kalman filtering

Ingeniería Superior de Telecomunicación

Bárbara Valenciano Martínez

Noviembre 2008

Speech enhancement using Kalman filtering

AUTOR: Bárbara Valenciano Martínez
TUTOR: Tom Bäckström
PONENTE: Doroteo Torre

Área de Tratamiento de Voz y Señales
Dpto. de Ingeniería Informática
Escuela Politécnica Superior
Universidad Autónoma de Madrid
Noviembre 2008

PROYECTO FIN DE CARRERA

Título: *Speech enhancement using Kalman filtering*

Autor: Bárbara Valenciano Martínez

Tutor: Tom Bäckström

Ponente: Doroteo Torre Toledano

Tribunal:

Presidente: Joaquín González Rodríguez

Vocal: Miguel Ángel García García

Vocal secretario: Doroteo Torre Toledano

Fecha de lectura:

Calificación:

Abstract:

This project studies the enhancement of the speech on telephonic conversations using Kalman filtering to remove the background noise.
For this, the database called AURORA (Aurora 4a) is used, which contains speech files with artificial addition of noise over a range of signal to noise ratios.

On the first part of this project we assume white noise and on the second one colored noise, carrying out the same experiments for both options.

The signal corrupted with noise is modeled through two coefficients prediction methods:

- Linear Predictive Coding (LPC)
- Stabilised weighted linear prediction (SWLP)

Finally it is used the Kalman filtering to filter the signal.

Keywords:
Speech enhancement, speech, Kalman filter, LPC, SWLP, state – space notation.

Resumen:

Este proyecto estudia la mejora de voz en conversaciones telefónicas utilizando filtrado de Kalman para la eliminación del ruido de fondo. Para ello hacemos uso de la base de datos AURORA (Aurora 4a) donde disponemos de ficheros de voz a los que se les ha añadido ruido de manera artificial sobre un rango de diferentes relaciones de señal a ruido.

En una primera parte del proyecto suponemos ruido blanco y en la segunda parte ruido coloreado, llevando a cabo los mismos experimentos para ambas opciones.

La señal dañada con ruido es modelada a través de dos métodos de predicción de coeficientes:

- Linear Predictive Coding (LPC)
- Stabilised weighted linear prediction (SWLP)

Por último realizamos un filtrado de Kalman de la señal.

Palabras clave:
Mejora del habla, habla, filtro de Kalman, LPC, SWLP, notación estado-espacio.

Agradecimientos

En primer lugar me gustaría agradecer a mi tutor Tom Bäckström, por todo el apoyo que me ha brindado y por la oportunidad que me ofreció al poder desarrollar este proyecto en Helsinki University of Technology, así como a todo el grupo de Laboratory of Acoustics and Audio Signal Processing, por la acogida tan buena que me dieron.

También quiero agradecer a Doroteo Torre, porque gracias a él pude desarrollar mi proyecto desde otra universidad y desde España siempre me ayudó con todas las dudas que pudieron surgir.
Gracias a la Universidad Autónoma de Madrid y a todos los profesores que durante estos años me han ayudado a formarme en mis estudios, por sus consejos y experiencia.

Como no decir algo de mis compañeros de la universidad, gracias a ellos estos años han pasado como un suspiro y me quedo con grandes y buenos momentos. Gracias a Cris que ha estado ahí para lo bueno y lo malo, por todas las largas noches de prácticas y todas las largas noches de fiesta. Gracias a Lucas por todo el apoyo, las conversaciones y por compartir conmigo su cultura musical. Gracias a Abejón y a Andrés porque nunca han dudado un segundo a la hora de echar una mano. Gracias a Nacho por su paciencia y porque sin él este proyecto no habría visto la luz. Y gracias en general a todos los que han hecho de mi estancia en la universidad una experiencia inolvidable: Tupi, David, Alfonso, Peter y Fer Alonso.

El Erasmus fue algo increíble y en gran parte fue debido a la gente que conocí allí. En primer lugar gracias a Vero, porque simplemente por conocerla y compartir esa experiencia con ella ya mereció la pena cruzar Europa, y en segundo lugar gracias a todos mis amigos de Leppävaara, en especial a mis compis de piso Cris y Ula. A cada uno de ellos le pertenece un trocito de este proyecto.

Gracias a Ana, a Mariu y a todos mis amigos de fuera de la universidad, simplemente por estar ahí y quererme un montón.

Y por último, pero no menos importante, gracias a mis padres y mis hermanas, por haber llegado conmigo hasta aquí, desde el primer día de cole hasta hoy, por el cariño, los consejos, el apoyo y la confianza.

Bárbara Valenciano Martínez
Noviembre 2008.

This Master Thesis has been developed in the Laboratory of Acoustics and Audio Signal Processing belonging to the Electrical and Communications Engineering Department, Helsinki University of Technology.

Table of contents

Table of figures

Introduction

Motivations and goals

Nowadays, everybody lives connected to a phone. It is of vital importance to be able to always communicate with others wherever we are. This means that, in most of situations, the place where we are doesn’t satisfy the most appropriate requirements to have a conversation.

Once that we have beaten the coverage problem, (lately it is possible to speak even from the subway through a mobile phone), appears another difficulty: the background noise. It is not useful to keep the communication if the noise of the subway, train, car, etc., prevents us from understanding.

That is the goal of this master thesis: we treat to eliminate the maximum background noise keeping the quality of the voice we want to transmit, because there exists a limit where the voice stops to sound like human voice, it loses the nuances and starts to seem like something artificial.

For this purpose we use the Kalman filter, which receives its name because of its researcher Rudolf Kalman, who based his studies on the Wiener filter.

For the experimental part we use the database AURORA Project Database (Aurora 4a), where we find sentences recorded in English, which have been filtered by IRS filters to simulate the telephone handsets and later, it has been added the background noise.

We will develop this project in two environments, with white noise and with colored noise; the latter is closer to the real life noise. For each scenario, two methods will be used to model the signals: LPC and SWLP. Both predict the coefficients of the signals.

At the end of the document, in the results section, we will compare all the scenarios and methods through two kinds of graphs, spectrogram and SNR method.

Motivaciones y objetivos

Hoy en día, todo el mundo vive conectado a un teléfono. Es de vital importancia estar comunicado con el resto del mundo donde quiera que nos encontremos. Esto significa que, en la mayoría de las situaciones, el lugar en el que nos encontramos no satisface los requerimientos más apropiados para llevar a cabo una conversación.

Una vez superado el problema de la cobertura, (últimamente es posible incluso hablar desde el metro con un móvil), aparece otra dificultad: el ruido de fondo. No es útil mantener una conversación si el ruido del metro, tren, coche, etc., no nos permite entendernos.

Ése es el objetivo de este proyecto: tratar de eliminar el máximo ruido de fondo para mantener la calidad de la voz que queremos transmitir, porque existe un límite donde la voz deja de parecer humana, pierde sus matices y empieza a parecerse a algo artificial.

Para este fin utilizamos el filtro de Kalman, el cual recibe su nombre a su investigador Rudolf Kalman, que basó sus estudios en el filtro de Wiener.

Para la parte experimental utilizamos la base de datos AURORA Project (Aurora 4a), donde se encuentran frases grabadas en inglés, las cuales han sido filtradas con un filtro IRS para simular las características telefónicas y después, se ha añadido el ruido de fondo.

Desarrollaremos este proyecto para dos situaciones, con ruido blanco y con ruido coloreado; este último es más parecido al ruido de la vida real. Para cada escenario, utilizaremos dos métodos para modelar las señales: LPC y SWLP. Ambos predicen los coeficientes de las señales.

Para finalizar el documento, en la sección de resultados, compararemos todos los escenarios y métodos utilizando dos tipos de gráficos: espectrogramas y representaciones de los niveles de SNR.

Document structure

This document has the following sections:

- State of the art about speech, going through different concepts like speech perception, communications, environment, etc. The last section of this one is about speech enhancement, it is the most important because our project is focused on it.
- Kalman filtering, in this section we can see everything about the Kalman filter, the algorithm and the mathematic background taking into account the advantages and the disadvantages of this type of filtering.
- Design and development. This is a description of our project, the material used, the database and the code developed.
- Results of the research.
- Conclusions of the work carried out and future work.
- References
- Appendix

State of the art

1. Speech

Speech is the process associated with the production and perception of the noises used in the spoken language. A huge number of disciplines study the speech and the speech sounds, including acoustic, psychology, speech pathology, linguistic, cognitive science and computer science.

Spoken language is used to communicate information from a speaker to a listener. Speech production and perception are both important components of the speech chain.

1.1 Speech perception

Speech perception refers to processes by which humans are able to interpret and understand the sounds used in the language. The study of the speech perception is closely linked to the phonetic field and phonology. Speech perception researches seek to understand how the humans recognize the speech sounds and use this information to understand the spoken language. The researches about the speech have applications in the building of computer systems which can recognize the speech, as well as improve the recognition for hearing impaired listeners.

There are a lot of biological and psychological factors which can affect the speech: disorders with the lungs, vocal cords, respiratory affections among others.

1.2 Speech communications

Speech is the most primary human communication. For that reason, it exists a big trend to increase and improve telecommunications. Nowadays, all the people use the communication devices almost as a primary good: telephones, mobiles, internet…and the customers demand a high coverage and quality.
However, the background noise is an important handicap. If it is joined with other distortions, it can seriously damage the service quality.
Added to this human-human interaction, it also exists a human-machine interaction based on a graphical user interface. However, still today the computers have a lack of human abilities like speaking, listening, understanding and learning.

As J.Benesty (2005) says: We live in a noisy world! In all applications (telecommunications, hands-free communications, recording, human-machine interfaces, etc) that require at least one microphone, the signal of interest is usually contaminated by background noise and reverberation. As a result, the microphone signal has to be "cleaned" with digital signal processing tools before it is played out, transmitted, or stored.

Speech processing is the study of speech signals and the processing methods of these signals. The signals are usually processed in a digital representation whereby speech processing can be seen as the intersection of digital signal processing and natural language processing. Speech processing can be divided in the following categories:

- Speech recognition, which deals with analysis of the linguistic content of a speech signal.
- Speaker recognition, where the aim is to recognize the identity of the speaker.
- Enhancement of speech signals (this is the area of this project)
- Speech coding, a specialized form of data compression, which is important in the telecommunication area.
- Voice analysis for medical purposes, such as analysis of vocal loading and dysfunction of the vocal cords.
- Speech synthesis: the artificial synthesis of speech, which usually means computer generated speech.
- Speech enhancement: enhancing of the perceptual quality of speech signal by removing the destructive effects of noise, limited capacity recording equipment, impairments, etc.

The speech processing has a lot of applications; one of them could be a tickets sales system by phone, where, without the necessity of an operator, a customer can buy tickets with different characteristics and options thanks to the words recognition systems.

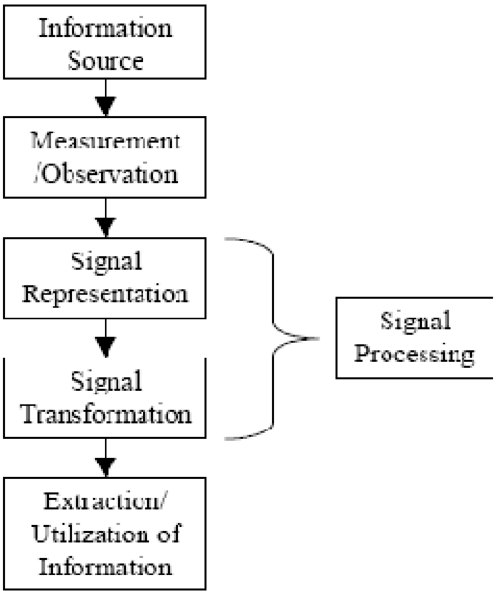


Figure 1.1: Speech processing

Figure 1.1 is a representation of the speech that ensures that the information content can be easily extracted by human customers or computers.

1.3 The acoustical environment

The acoustical environment is defined as a set of transformation that affect the speech signal since the moment it leaves the speaker's mouth until it is in digital form. There are, among others, two main sources of distortion: additive noise and channel distortion:

- Additive noise is like a fan running in the background, a door slam, a conversation among others; it is in our common daily life. It can be stationary or non stationary.
  Stationary noise is the one made by a computer fan or air conditioning; it has a spectral power density that does not change over time.
  Non stationary noise, caused by door slam, radio, TV, voices, has statistical properties that change over time. A signal captured with the speaker close to the microphone has a little noise and reverberation. However, if the microphone is far from the speaker's mouth it can pick up a lot of noise and/or reverberation.
- Channel distortion can be caused by reverberation, the frequency response of a microphone, the presence of an electrical filter in an A/D circuit, the response of the local loop of a telephone line, a speech codec, etc. Reverberation, caused by the reflection of acoustic waves on the walls and other objects, can also dramatically alter the speech signal.

If both the microphone and the speaker are in an anechoic chamber or in free space, the microphone picks up only the direct acoustic path. In the practice, in addition to the direct acoustic path, there are reverberations from the walls and other objects in the room. The signal level at the microphone is inversely proportional to the distance from the speaker, for the direct path. For the reflected sound waves, the sound has to travel a larger distance, and its signal level is proportionally lower. Moreover, we have to take into account the energy absorption which takes place each time the sound wave hits a surface.

1.4 Speech enhancement

What is speech enhancement? Enhancement means the improvement in the value or quality of something. When applied to speech, this simply means the improvement in intelligibility and/or quality of a degraded speech signal by using signal processing tools. By speech enhancement, it refers not only to noise reduction but also to dereverberation and separation of independent signals. Since this field is fundamental for research in the applications of digital signal processing, it is also of great interest to the industry which is always looking for new solutions that are both effective and practical.

This is a very difficult problem for two reasons. First, the nature and characteristics of the noise signals can change dramatically in time and between applications. It is also difficult to find algorithms that really work in different practical environments. Second, the performance measure can also be defined differently for each application. Two criteria are often used to measure the performance: quality and intelligibility. It is very hard to satisfy both at the same time.

Speech enhancement is an area of speech processing where the goal is to improve the intelligibility and/or pleasantness of a speech signal. The most common approach in speech enhancement is noise removal, where we, by estimation of noise characteristics, can cancel noise components and retain only the clean speech signal. The basic problem with this approach is that if we remove those parts of the signal that resemble noise, we are also bounded to remove those parts of the speech signal that resemble noise. In other words, speech enhancement procedures, often inadvertently, also corrupt the speech signal when attempting to remove noise. Algorithms must therefore compromise between effectiveness of noise removal and level of distortion in the speech signal.

Current speech processing algorithms can roughly be divided into three domains, spectral subtraction, sub-space analysis and filtering algorithms:

- Spectral subtraction algorithms operate in the spectral domain by removing, from each spectral band, that amount of energy which corresponds to the noise contribution. While spectral subtraction is effective in estimating the spectral magnitude of the speech signal, the phase of the original signal is not retained, which produces a clearly audible distortion known as "ringing".
- Sub-space analysis operates in the autocorrelation domain, where the speech and noise components can be assumed to be orthogonal, whereby their contributions can be readily separated. Unfortunately, finding the orthogonal components is computationally expensive. Moreover, the orthogonality assumption is difficult to motivate.
- Finally, filtering algorithms are time-domain methods that attempt to either remove the noise component (Wiener filtering) or estimate the noise and speech components by a filtering approach (Kalman filtering).

There is an important algorithm for speech enhancement which belongs to the group of parametric methods where the speech signal is modeled as an autoregressive process embedded in Gaussian noise. Speech enhancement algorithms belonging to this category consist of two steps:

- Estimation of the AR coefficients and noise variances.
- Application of the Kalman filtering using the estimated parameters to estimate the clean speech from a sample of the noisy signal.

After this, instead of using linear prediction like Gannot did, we will use Stabilized Weighted Linear Prediction (SWLP).

2. Speech modeling

The modeling of speech studies how humans produce the voice. Nowadays we have a lot of devices which "speak" to us and this voice should be as similar as possible to a real human voice. For that reason, a lot of researches are aimed to find a good model of speech production (Figure 2.1).

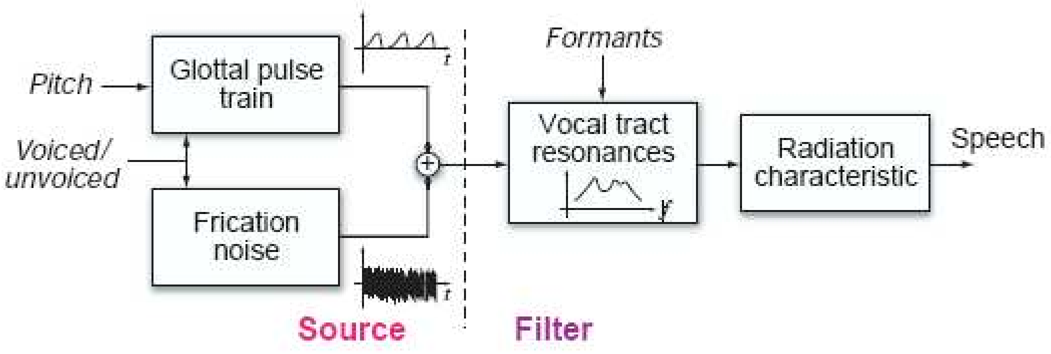


Figure 2.1: Production model voice

First of all, with this model we decide if the noise that we want to produce is voiced or unvoiced.
For the voiced sounds we have to model a glottal pulse train similar to the produced in our vocal cords. For the unvoiced sounds the signal produced is like noise, similar to the signal that we can see in the fricative sounds.

After that we have to go through the vocal tract with our generated signal. In this section we filter the signal with a filter that tries to imitate the effect of the shape formed with the pharyngeal cavity (throat), vocal and nasal cavity.

Finally the radiation model reproduces the effect of the radiation impedance that the air put up to the exit of the speech from the mouth.

2.1 Speech production mechanism

The main components of the speech system (Figure 2.2) are the lungs, trachea, larynx (organ of the speech production), pharyngeal cavity (throat), vocal or oral cavity (mouth) and nasal cavity (nose). In techniques discussions, the oral cavity and pharyngeal are grouped in one unit which is referenced as vocal tract, and the nasal cavity is sometimes called nasal tract. The vocal tract starts at the end of the larynx (vocal cords or glottis) and ends at the beginning of the lips. The nasal tract starts with the velum and finishes with the nostrils. When the velum goes down, the nasal tract is acoustically coupled to the vocal tract to produce the speech nasal sounds.

When we breathe the air comes from the lungs and then it is expelled from them through the trachea making the vocal cords to vibrate.
The air is divided in quasi periodic pulses which are modulated in frequency when they cross the throat, the oral cavity or maybe the nasal one. Depending on the position of the tongue, jaw, teeth, lips, etc. we produce different sounds.

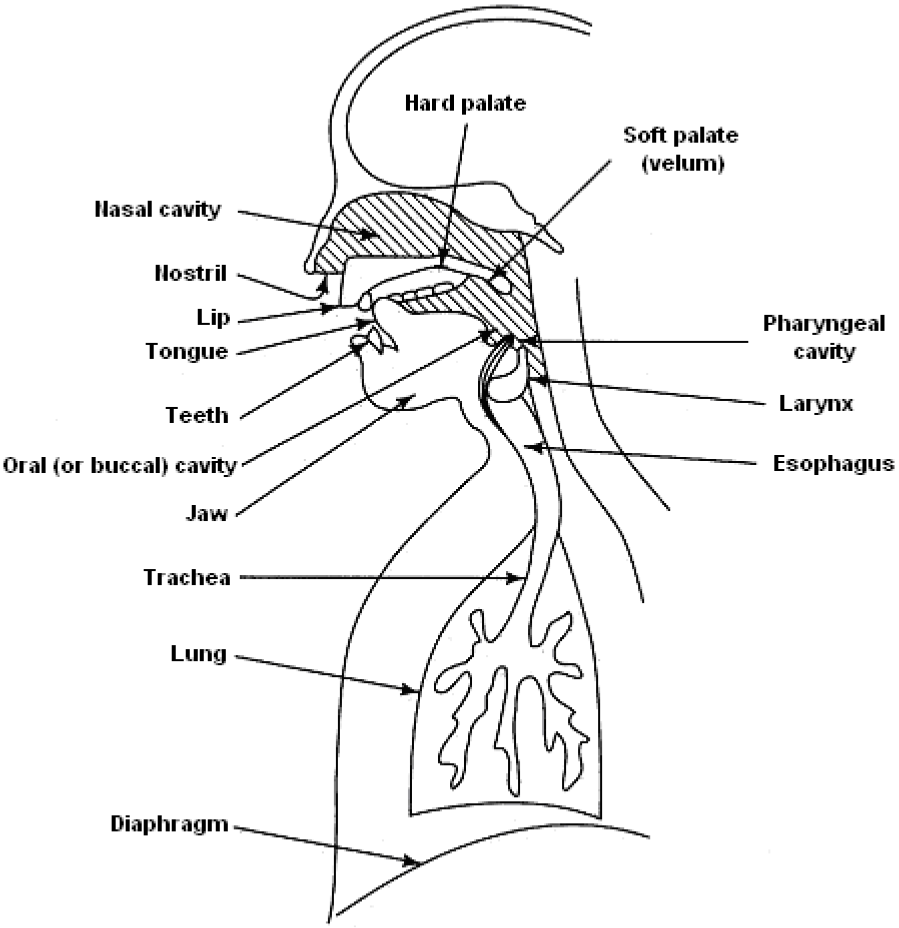


Figure 2.2: Speech production organs

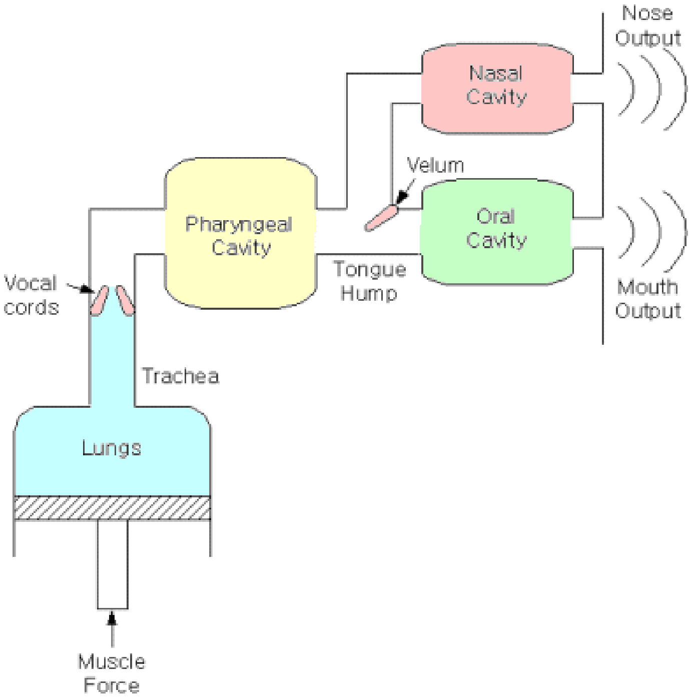


Figure 2.3: Simplified model for speech production

Humans use the language almost unconsciously, without paying attention in how the information is processed by the brain, the amount of organs involved in this process or the different phases in this communication.

Speech begins with a thought and an intention to communicate in the brain, which activates the muscular movements to produce speech sounds. A listener receives a sound in the auditory system, processing it for a conversion to neurological signals the brain can understand. The speaker continuously monitors and controls the vocal organs by receiving his or her own speech as feedback.

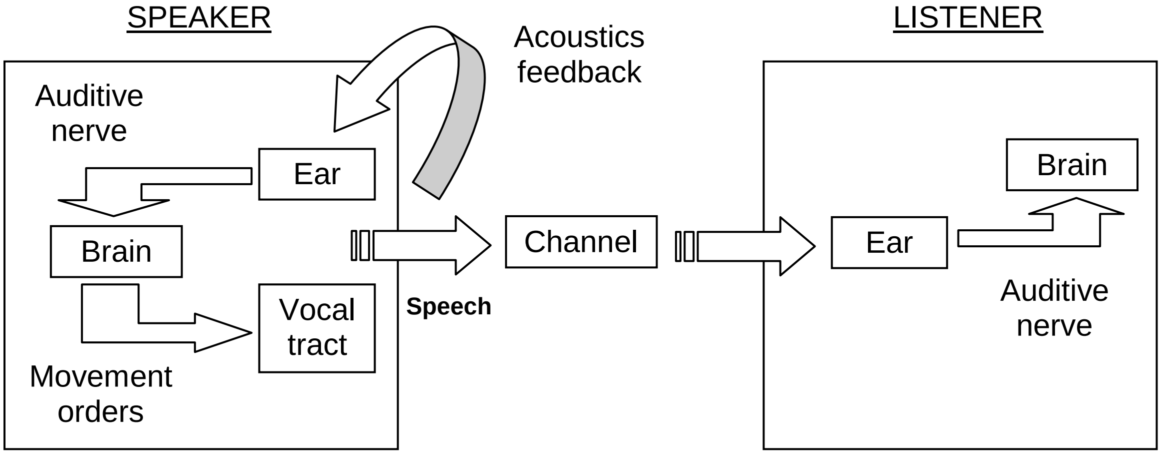


Figure 2.4: Speech generation and speech understanding

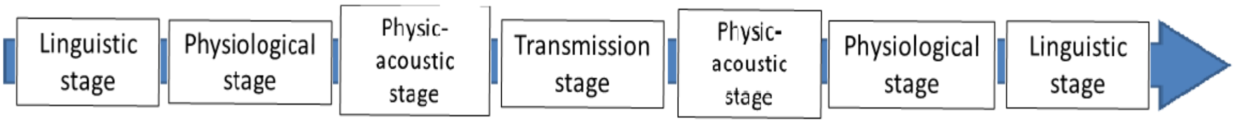


Figure 2.5: Stages in the spoken communication

All this activity starts in the speaker's mind with a message to be transmitted to the listener via speech. This is the linguistic stage.
After a message is created, the next step is to convert the message into a sequence of words. Each word consists of a sequence of phonemes that correspond to the pronunciation of the words.
The spoken signal appears when the air crosses the trachea from the lungs. This air crosses the vocal cords, situated in the larynx, which have two functions:

- With the voiced sounds the vocal cords are in tension and they vibrate when the air goes across them.
- With the unvoiced sounds the vocal cords are relaxed and the air can cross them freely.

In the next step, the brain sends the information to the vocal tract, where the air takes the characteristics of each formant. This is the physiological stage.
After that, when the speaker starts to speak we are in the physic-acoustic stage. The sounds are materialized but, at the same time, there is a feedback because the ear of the speaker can hear what is he saying and the brain analyzes the meaning. And the process in the speaker's brain that starts is the same that when it is the listener who speaks and the speaker who listens.
The sounds travel by the air in the transmission stage from the mouth of the speaker to the ear of the listener. In this stage, when the sounds cross the channel, the noise and other distortions are added.

Finally, we have the same stages but in the listener side in reverse order. First the message is passed to the cochlea in the inner ear, which performs frequency analysis as a filter bank. A neural transduction converts the spectral signal into activity signal on the auditory nerve.
Currently, it is unclear how the neural activity is converted into the language system and how the brain can achieve the comprehension of the message.

Speech signals are composed by analog units, which are the symbolic representation of the spoken language: phonemes, syllables and words. But for this project it is not necessary to go into this level of depth.

2.2 Tube model

The sounds are vibrations transmitted by the air. Due to the fact that they are waves, all the physics laws describe their behavior. The solution of the equations which describe these waves is very difficult to find, unless we assume some simplifications for the vocal tract, and the energy losses. As Rabiner and Schafer (1978) say: the most simple configuration of the vocal tract is to model it as a non-uniform tube, time-varying, cross-section assuming that there are not losses due to the viscosity, bulk of the fluid or walls of the tube.

If, otherwise, we want a detailed study we have to consider:

1. Time variation of the vocal tract shape
2. Losses due to heat conduction and viscous friction at the vocal tract walls
3. Softness of the vocal tract walls
4. Radiation of sound at the lips
5. Nasal coupling
6. Excitation of sound in the vocal tract

To find a solution and to model these sound waves we have to consider some conditions to study them:

- The limit conditions in each tube end:
  - Lip: consider the effects of the sound radiation
  - Glottis: know the source of the excitation
- The area function of the vocal tract A(x,t)

When we emit a constant sound, the vocal tract area does not change. However, when we speak this area changes constantly making more difficult its study. There are a lot of methods to observe and study this area changes but even with them, the solution of the sound waves is still very complicated. Fortunately, we can achieve a possible solution with some approximations.

The simplest way is to study the vocal tract simplifying it as a uniform lossless tube. If we want to make this model more realistic we can concatenate more than one uniform lossless tube, each one with appropriate characteristics, and considering each one as ideal transmission lines.

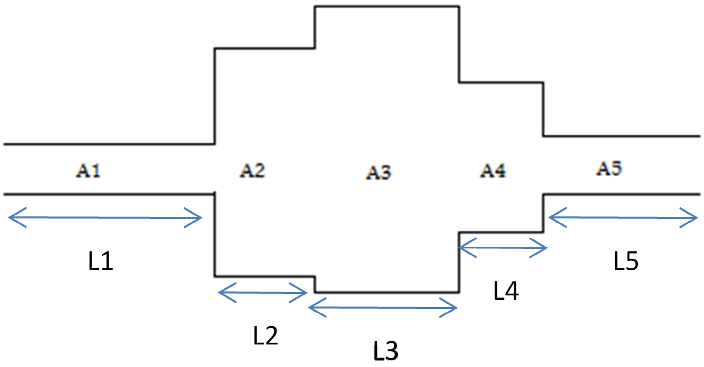


Figure 2.6: Simulation of vocal tract by concatenation of uniform lossless tubes

These concatenations of tubes would make the model closer to the reality. The area of each section is chosen in order to approximate it to the area function A(x) of the vocal tract. If we use a large number of tubes, but with shorter length, our model would be closer to our real goal, because we would represent better how it changes the area function. In this way, we have to group all the losses of the vocal tract at the lips and glottis.

If we consider the effects of the losses in the vocal tract, the model is more complicated due to the fact we have to consider the sound waves friction with the air and with the tube walls. Furthermore, in the real model, the walls of the vocal tract are in continuous movement because the air that flows inside them changes its pressure each moment, varying, in this way, the vocal tract area.

Other phenomenon that we should consider if we want to study the real model is the radiation effect in the lips and the diffraction effect that is caused.

2.3 Linear Prediction (LP)

Linear prediction is one of the most powerful tools used, where a signal y_n is the output of a system considering the unknown signal x_n as the input with the relation,

$$y_n = -\sum_{k=1}^{p} a_k \cdot y_{n-k} + G \cdot \sum_{l=0}^{q} b_l \cdot x_{n-l} \quad , \quad b_0=1$$

where G are the parameters of a hypothesized system.
In this equation we can see that the output y_n is a linear combination of the past samples at the system exit and past and present inputs. The name Linear Prediction comes from this formula, which shows that a

signal y_n can be predicted from linear combinations of past and present outputs and inputs.
This equation can be also written in the frequency domain, taking the z transform on both sides. If H(z) is the transfer function of the system, then we have:

$$H(z) = \frac{Y(z)}{X(z)} = G \cdot \frac{1 + \sum_{l=1}^{q} b_l \cdot z^{-l}}{1 + \sum_{k=1}^{p} a_k \cdot z^{-k}} \qquad \text{where,} \qquad S(z) = \sum_{n=-\infty}^{\infty} y_n \cdot z^{-n}$$

The roots of the numerator and denominator polynomials of H(z) are the zeros and poles of the model, respectively. There are two special cases of the model:

- All-zero model: $a_k = 0$, $1 \leq k \leq p$ known as moving average (MA) model
- All-pole model: $b_l = 0$, $1 \leq l \leq q$ known as autoregressive (AR) model

The estimation of model parameters can be derived in the time domain and in the frequency domain.

In general, we don't know the input signal x_n, we have to predict it, $\hat{x}_n$, as a linear weight combination of the past samples,

$$\hat{y}_n = -\sum_{k=1}^{p} a_k \cdot y_{n-k}$$

where a_k are the predictor coefficients, p is the model order and the minus sign is for convenience. If we want to know the error with this method,

$$e_n = y_n - \hat{y}_n = y_n + \sum_{k=1}^{p} a_k \cdot y_{n-k} = \sum_{k=0}^{p} a_k \cdot y_{n-k}$$

where y_n is the original signal, a_0=1 and e_n is called residual.

The idea is to get an error as small, close to zero, as possible; this measures the quality of the predictor. If we denote the total squared error by E, where

$$E = \sum_n e_n^2 = \sum_n \left(y_n + \sum_{k=1}^{p} a_k \cdot y_{n-k} \right)^2$$

Then, to minimize E:

$$\frac{\partial E}{\partial a_i} = 0, \ 1 \leq i \leq p$$

This method is called method of least squares and the parameters a$_k$ are calculated as a result of the minimization of the mean or total squared error with respect to each of the parameters (it is called autocorrelation criterion too).

For all the definitions of y$_n$, we can find a set of p equations with p unknowns and solve them for the predictor coefficients which minimize E, with the autocorrelation method,

$$\sum_{k=1}^{p} a_k \cdot R(i-k) = -R(i), \quad 1 \le i \le p$$

$$E_p = R(0) + \sum_{k=1}^{p} a_k \cdot R(k)$$

where,

$$R(i) = \sum_{n=-\infty}^{\infty} y_n \cdot y_{n+1}$$

is the autocorrelation function of y$_n$. Then, we can observe that:

$$R(-i) = R(i)$$

Like the coefficients R(i-k) form an autocorrelation matrix, this method is called, as we said above, autocorrelation method. An autocorrelation matrix is a symmetric Toeplitz matrix (where all the elements of each diagonal are equal). The development of the equation would be:

$$\begin{bmatrix} R_0 & R_1 & R_2 & \cdots & R_{p-1} \\ R_1 & R_0 & R_1 & \cdots & R_{p-2} \\ R_2 & R_1 & R_0 & \cdots & R_{p-3} \\ \vdots & \vdots & \vdots & & \vdots \\ R_{p-1} & R_{p-2} & R_{p-3} & \cdots & R_0 \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_p \end{bmatrix} = -\begin{bmatrix} R_1 \\ R_2 \\ R_3 \\ \vdots \\ R_p \end{bmatrix}$$

2.3.1 <u>Linear Predictive Coding (LPC)</u>

Linear predictive coding (LPC) is a tool used, mainly, in the audio signal and speech processing to represent the spectral envelop of a speech digital signal in a compressed way (using the information of linear prediction model). This technique is one of the most powerful to analyze the speech, and one of the most useful methods for encoding with good quality at low rate.

LPC starts with the assumption that the speech signal is produced by a buzz at the end of a tube, adding, sometimes, hissing and popping sounds. This model is a good approximation to the reality.

The glottis produces the buzz, which is characterized by his intensity (loudness) and frequency (pitch). The vocal tract generates a tube which is characterized by his resonances, called formants. The lips, tongue and throat generate the hisses and pops sounds.
LPC analyzes the speech signal using the formants, removing their effect from the speech signal and estimating the intensity and frequency of the remaining speech signal buzz. The removing formants process is called inverse filtering and the remaining signal after the subtraction is called residue.
The numbers which describe the frequency and intensity of the buzz, the formants and the residue signal can be stored or transmitted.

LPC synthesizes the speech signal with the inverse process: it uses the buzz and residue parameters to create a source signal and after that it uses the formants to create a filter (which represents the tube) and then runs the source signal through the filter we get the speech.
Due to the fact that speech signals change over the time, this process is realized with small chunks of speech signal, called frames. Usually with a number between 30 and 50 frames per second we get a speech signal intelligible and a good compression.

The basic problem of the LPC system is to determine the formants from the original signal. The solution is to express each sample as a linear combination of previous samples. This equation is called linear predictor. The coefficients of the equation (the prediction coefficients) characterize the formants, so we use the LPC system to estimate these coefficients.

2.4 Spectral models

2.4.1 Improved spectral models: LP, WLP, MVDR

LP is the most common method to speech modeling, but it has some disadvantages:

- Such as the biasing of the formant estimates by their neighbouring harmonics.
- The effectiveness is lower in presence of noise.

For that reason, some methods of linear prediction have been developed with a better robustness against the noise. Most of these improvements are based on the iterative update of the predicted parameters.

The first of them is weighted linear prediction (WLP) which tries to confront the problem caused by the glottal closure excitation by introducing an energy weight in the time domain of the error prediction. As it stresses those segments which have a high SNR, WLP improves the spectral envelopes in noisy conversations (if we compare it with LP).

Furthermore, the filters of WLP can be calculated without iterative updates.

The second one is the minimum variance distortionless response method (MVDR) and it is popular in arrays processing, but lately it is becoming useful to make extractions in speech recognition.
Among other refinements studied in some researches, the scaling of spectral envelope improves the robustness of the MVDR spectral model against additive noise in the frequency domain.

2.4.2 Stabilised weighted linear prediction (SWLP)

All the information above takes us to compare the all-poles models: LP, WLP, and MVDR. Like the first version of WLP didn't guaranteed us the stability of the all-poles model, the idea was to improve it developing weighted functions which make the model stable. Here is where the SWLP method appears. Choosing correctly the parameters, we have similar envelopes to those obtained with MVDR method, but with an improved robustness against the background noise.

In our case, the idea is to find an optimization of the filter parameters in stabilized weighted linear prediction. For that we have to find the coefficient vector a = $(a_0\ a_1\ \ldots\ a_p)^T$, of a FIR predictor with order p, which minimizes the prediction error energy. The corresponding all-pole filter is obtained as H(z) = 1 / A(z), where A(z) is the z-transform of a.

To achieve this, it exists a formula to modify the weight function of WLP and, in this way, to reach the stability of the all-poles filter. All of this can be carried out by changing the elements of the secondary diagonal of the B matrix:

$$B_{i,i+1} = \begin{cases} \sqrt{w_i / w_{i+1}}\ , & \text{if } w_i \leq w_{i+1} \\ 1, & \text{if } w_i > w_{i+1} \end{cases}$$

From now on, the WLP method calculated using the B matrix, is called stabilized weighted linear prediction (SWLP), where the stability of the all-poles filter is guaranteed.

The main concept in WLP, is the time domain weight function. Choosing an appropriate waveform, one can temporally emphasize or attenuate the weight of the residual energy prior to the optimization of the filter parameters. The weight function was chosen basing on the short-time energy (STE),

$$w_n = \sum_{i=0}^{M-1} x_{n-i-1}^2$$

where M is the length of the window. The idea of the weight, is that computing linear predictive models of speech are more robust against the noise than the traditional LP. This is based on the fact that the STE function emphasizes those sections of the speech waveform which have samples of large amplitude. These segments of speech are less vulnerable to noise in comparison to those values with smaller amplitude.

Using the results of the article "Stabilised Weighted Linear Prediction- A Robust All-Pole Method for Speech Processing" (Magi et.al. 2007) we can observe the behavior of SWLP in spectral modeling of speech:

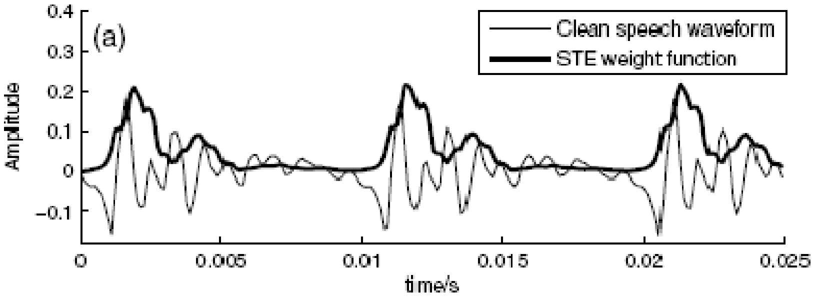


Figure 2.7: Time domain waveforms of clean speech and STE weight function (window M=8)

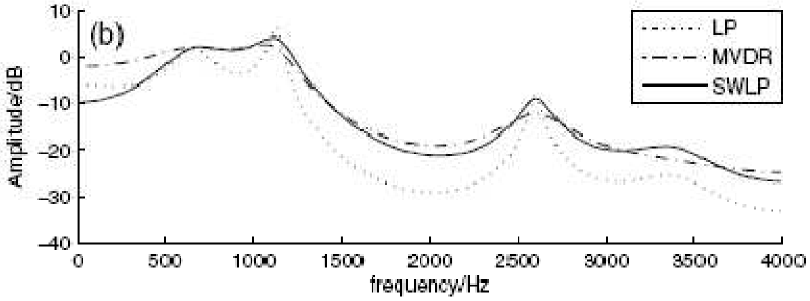


Figure 2.8: All-poles spectra of order p=10 computed by LP, MVDR and SWLP (from Figure 2.7)

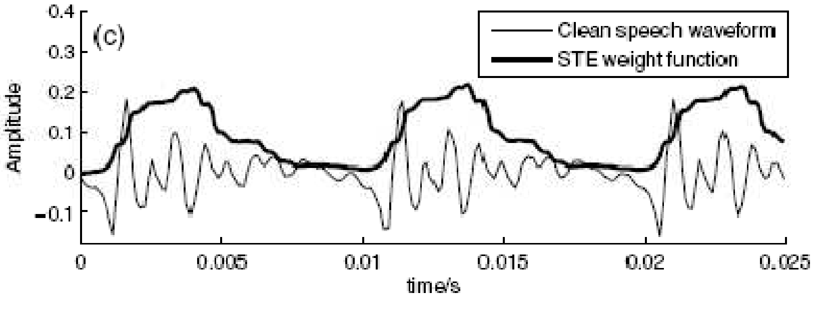


Figure 2.9: Time domain waveforms of clean speech and STE weight function (window M=24)

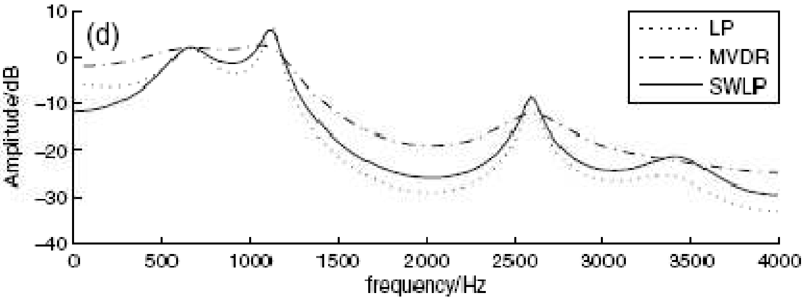


Figure 2.10: All-poles spectra of order p=10 computed by LP, MVDR and SWLP (from Figure 2.9)

In the Figures 2.7 and 2.9 we can see the analyzed speech sound with the STE weight function. Below each one of them, we can see the spectra of parametric all-poles models with order p=10 with the three techniques: LP, MVDR, SWLP.

To demonstrate the importance or the effect of the window's length, the SWLP model with M=8 (Figure 2.7) and M=24 (Figure 2.9) is analyzed.

At the time-domain panels we can see how the weight function calculated with STE emphasizes the segments which have higher amplitudes, while the segments with lower amplitudes are less emphasized.

At the spectra panels, we can appreciate how the SWLP spectrum changes its shape with the window's length. With M=8, the variations of the spectrum are very smooth, while with M=24 the spectrum is sharper.

In short-term, we can say that the filter with SWLP model with large values for M is more similar to the behavior with the LP model, and with small values for M we put nearer MVDR model.

Taking into account the experiment made in the article mentioned above, if we study how SWLP method works for speech corrupted by additive noise and after that we compare the performance to that of LP and MVDR, we can see that SWLP presents the best robustness against noise. With a small value for M, SWLP is able to face the effect of additive noise in a more effective way than the other methods.

3. Kalman filtering

The filter has its origin in a Kalman's document (1960) where it is described as a recursive solution for the linear filtering problem for discrete data. The research was in a wide context of state – space models, where the point is the estimation through the recursive least squares. Since that moment, due to the development of digital calculation, Kalman filter has been researched and applied, particularly in self and assisted navigation, missiles search and economy.

The study of Kalman filter is based on Wiener filter.

3.1 Wiener filter

This filter is the precursor of Kalman filter. The goal of Wiener filter is to remove the noise from a corrupted signal.

In general there are two processes which affect the signal that we want to measure:

- First of all, it is a fact that every device introduces an error in the output when a signal is measured. If our original signal is x_k and the response of the device is h_k our signal in the output is:

$$y_k = x_k * h_k \leftrightarrow Y_j = X_j \cdot H_j$$

- Secondly, the signal outside has noise added due to the process.

$$\hat{y}_k = y_k + n_k$$

To solve this equation, if we don't have noise and we know the response, then the solution is easy to find:

$$X_j = \frac{Y_j}{H_j}$$

But if we have noise, we have to filter the output signal with a Wiener filter.

$$X_j = \frac{Y_j \cdot W_j}{H_j}$$

For that, we should find the optimal Wiener filter. This kind of filter was proposed by Norbert Wiener during the 1940s. To reduce the amount of noise in the corrupted signal this filter is based on a statistical approach.

Normally, the filters are designed for a specific frequency, but in Wiener filters, first of all, we have to have knowledge about the spectral properties of the original signal and noise, and after that, we have to find a LTI filter whose output would be as close as possible to the original signal.

The Wiener filters are characterized by the following concepts:

- Assumption: signal and (additive) noise are stationary linear stochastic processes with known spectral characteristics or known autocorrelation and cross-correlation.
- Requirement: the filter must be physically realizable, i.e. causal (this requirement can be dropped, resulting in a non-causal solution).
- Performance criteria: minimum mean-square error.

3.2 Kalman filter

The filter is a mathematical procedure which operates through a prediction and correction mechanism. In essence, this algorithm predicts a new state from its previous estimation by adding a correction term proportional to the predicted error. In this way, this error is statistically minimized. This filter is the main algorithm to estimate dynamic systems specified in state-space form.

A Kalman filter is simply an optimal recursive data processing algorithm.

If we focus on the word optimal, its definition depends on the criteria chosen to evaluate. A feature is called optimum if the Kalman filter incorporates all the information provided. It processes all the measurements available, regardless the precision, to estimate the current value of the interest variables, using:

- Knowledge of the system and the measurement devices.
- Statistic description of the system noises, measurements of errors and the uncertainty of the dynamics models.
- Any information available about the initials conditions of the variables under study.

A Kalman filter would be built to combine all these data and with the knowledge of some dynamic systems to generate the best estimation of the interest variable.

If, on the contrary, we focus on the word recursive, this means that the Kalman filter doesn’t require storing all the previous samples and it neither needs to reprocess them on each new measurement taken. This feature is very important to the filter practicality.

We say that this is a data processing algorithm because it is just a computer program in a processing central.

The complete estimation procedure is as follows:
The model is formulated on state-space and for an initial set of parameters given, the model prediction errors are generated by the

filter. These are used recursively to evaluate the probability function until its maximization.

As a summary, we can say that the Kalman filter combines all the available data measured, plus the knowledge of the system and the measurement devices, to produce an estimation of the desired variables in such a manner that the error is statistically minimized.

3.2.1 **The discrete algorithm of Kalman filter**

The Kalman filter consists in a set of mathematic equations which give an optimum recursive solution through the least square method. The goal of this solution is to calculate an unbiased minimum variance linear estimator of the state in t, based on the information available in t-1, and update these estimations, with the additional information available in t, (Clar eh al. 1998). The filter is developed assuming the system can be described through a stochastic linear model, where the associated error to both, the system and the additional information which is incorporated on it, have a normal distribution with zero mean and a determinate variance.

The solution is optimum when the filter combines all the observed information and the previous knowledge about the system behavior to produce a state estimation so the error is statistically minimized. The recursive term means the filter recalculates the solution each time a new observation or measure is added to the system.

The Kalman filter is the main algorithm to estimate dynamics systems represented as state-space. In this representation the system is described by a set of variables denominated of state. The state contains all the information to do with a certain point in time. This information must permit the deduction of the past system behavior, with the goal of predicting its future behavior.

What makes the filter so interesting is its skill to predict the system's state in the past, present and future, although the nature of the system is unknown. In practice, the individual state variables of a dynamic system can't be determined exactly by a direct measure. Due to the foregoing, its measure is done with stochastic processes which have some uncertainty in the measure.

3.2.2 **The process to be estimated**

The Kalman filter has the goal of solving the general problem of estimate the state $X \in R^m$ of a process controlled in discrete time, which is dominated by a linear equation in stochastic difference in the following way:

$$X_n = A \cdot X_{n-1} + w_{n-1}$$

With a measure $Y \in R^n$, that is:

$$Y_n = C \cdot X_n + v_n$$

The random variables w_n and v_n represent the process and the measure error, respectively. It is assumed they are independent of each other, and are white noise variables with normal probability distribution:

$$p(w) \approx N(0,R_w)$$
$$p(v) \approx N(0,R_v)$$

In practice, the covariance matrix of the process's perturbation, R_w, and the measure's perturbation, R_v, could change in time, but for simplicity, it is assumed they are constants.

The matrix A is assumed to be of $m_x m$ dimension and it relates the state in the period n-1 with the state in the n moment. The matrix C has a dimension $n_x m$ and it relates the state with the measure Y_n. These matrixes may change over time, but generally they are assumed as constant.

3.2.3 The algorithm

The Kalman filter estimates the previous process using a feedback control, that is, it estimates the process to a moment over the time and then it gets the feedback through the observed data.

From the equation point of view that is used to derivate the Kalman filter, it is possible to separate them into two groups:

- Those which update the time or prediction equations
- Those which update the observed data or update equations

The first group of equations has to throw the state to the n moment taking as reference the state on n-1 moment and the intermediate update of the covariance matrix of the state. The second group of equations has to take care of the feedback; they add new information inside the previous estimation to achieve an improved estimation of the state.

The equations which update the time can be seen as prediction equations, while the equations which add new information can be seen as correction equations. Exactly, the final estimation algorithm can be defined as a prediction-correction algorithm to solve many problems. In this way, the Kalman filter works through a projection and correction mechanism to predict the new state and its uncertainty and correct the projection with the new measure. This cycle is showed in the following figure.

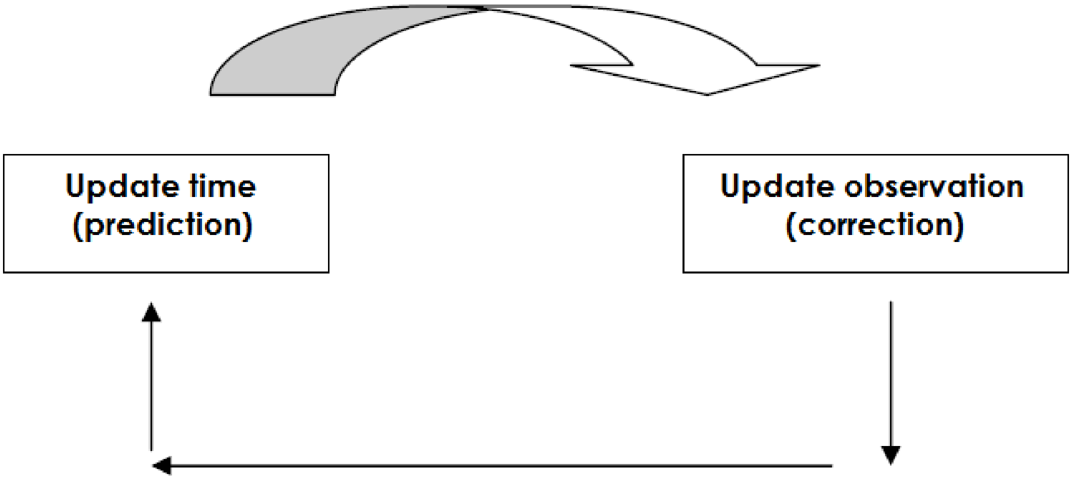


Figure 3.1: The Kalman filter cycle

The first step is to generate a state prognostic forward over the time taking into account all the information available at that moment, and the second step is to generate an improved state prognostic, so the error is statistically minimized.

The specified equations for the state prediction are detailed as follows:

$$x_{n|n-1} = A_{n,n-1} \cdot x_{n-1}$$

$$R_{e,n|n-1} = A_{n,n-1} \cdot R_{e,n-1} \cdot A^{T}_{n,n-1} + u \cdot R_{w} \cdot u^{T}$$

Notice how the equations predict the state and covariance estimations forward from moment n-1 to n.
These two formulas give us an estimate value for x_n and its covariance, when we don't have the real sample yet available.

The first Kalman equation estimates the next sample from the previous state. The second Kalman equation is the covariance matrix used to predict the estimation error. The A matrix relates the state in the previous moment n-1 with the actual moment n, this matrix could change for the different moments over the time. R_w represents the covariance of the process random perturbation which tries to estimate the state.

The specified equations for the state correction are detailed as follows:

$$R_{e,n} = R_{e,n|n-1} - R_{e,n|n-1} \cdot C^{T} \cdot F_{n}^{-1} \cdot C \cdot R_{e,n|n-1}$$

$$x_{n} = x_{n|n-1} + R_{e,n|n-1} \cdot C^{T} \cdot F_{n}^{-1}(y_{n} - C \cdot x_{n|n-1})$$

$$F_{n} = C \cdot R_{e,n|n-1} \cdot C^{T} + R_{v}$$

These are used when we have the real sample y_n. For that reason, they are called updating equations too.

The first task during the state projection correction is the calculation of the Kalman gain, $R_{e,n}$. This gain factor is chosen in such a way it minimizes the covariance error of the new state estimation.
The next step is to measure the process to get y_n and generate a new state estimation which incorporates the new observation.
The final step is to find a new estimation of the error covariance through the last equation.

After each couple of updates, time and measure, the process is repeated taking as starting point the new state estimations and the error covariance. This recursive nature is one of the most famous characteristics of Kalman filter.

The next figure offers us the complete operation of the filter, combining the previous figure and the five Kalman equations.

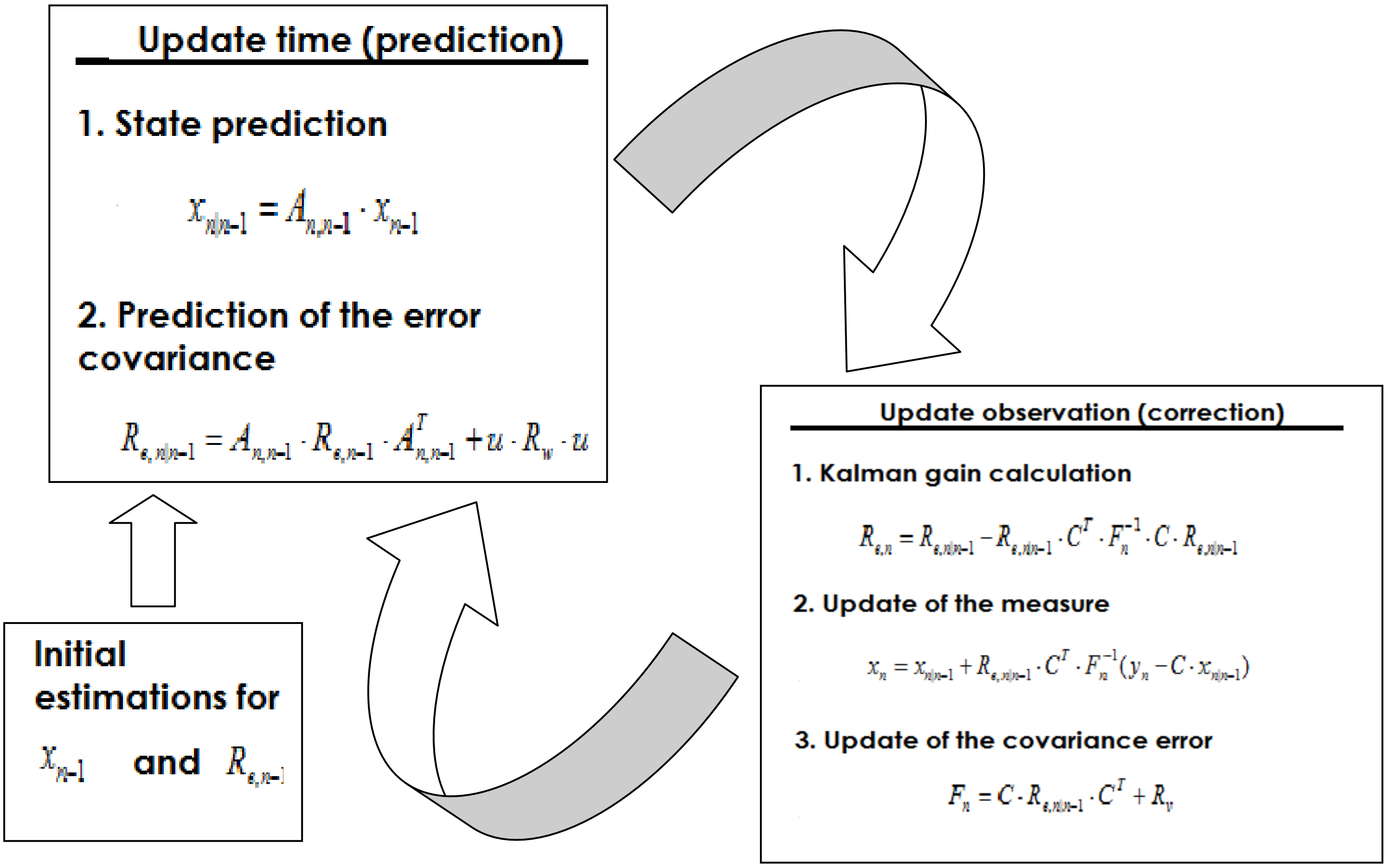

Update time (prediction)
1. State prediction
$x_{n|n-1} = A_{n,n-1} \cdot x_{n-1}$
2. Prediction of the error covariance
$R_{\varepsilon,n|n-1} = A_{n,n-1} \cdot R_{\varepsilon,n-1} \cdot A^{T}_{n,n-1} + u \cdot R_{w} \cdot u$
Initial estimations for
x_{n-1} and $R_{\varepsilon,n-1}$
Update observation (correction)
1. Kalman gain calculation
$R_{\varepsilon,n} = R_{\varepsilon,n|n-1} - R_{\varepsilon,n|n-1} \cdot C^{T} \cdot F_{n}^{-1} \cdot C \cdot R_{\varepsilon,n|n-1}$
2. Update of the measure
$x_{n} = x_{n|n-1} + R_{\varepsilon,n|n-1} \cdot C^{T} \cdot F_{n}^{-1}(y_{n} - C \cdot x_{n|n-1})$
3. Update of the covariance error
$F_{n} = C \cdot R_{\varepsilon,n|n-1} \cdot C^{T} + R_{v}$

Figure 3.2: Complete vision of Kalman filter. All these five equations make the Kalman filtering process

3.2.4 The Kalman filter and the state-space notation

The Kalman filter is the main algorithm to estimate dynamic systems specified with the state-space model. Actually, the state-space models and the Kalman filter models are often used as synonymous. The estimation and control of the problems of this methodology are based on stochastic models, assuming errors in the measures.

The performance of the state-space model for a linear system captures a y_n vector with nx1 order associated to an unknown x_n vector with mx1 order, known as state vector.

In speech processing, we assume the case with a signal received by a single microphone and additive noise. Let the signal measured by the microphone be given by:

$$y_n = x_n + v_n$$

Where yn is the observed signal, xn is the desired input and v_n is the additive background noise (zero-mean noise). Furthermore, like x_n is modeled as autoregressive, we assume the standard LPC modeling for the speech signal over an analysis frame:

$$x_n = -\sum_{k=1}^{m} a_k \cdot x_{n-k} + w_n = -a_n * x_{n-1} + w_n$$

On the other hand, the last equation can be reformulated in a state-space presentation with the state transition matrix or companion matrix:

$$A_{n+1,n} = \begin{bmatrix} -a_1 & -a_2 & \cdots & -a_{m-1} & -a_m \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix}$$

Then, we can write the state-space form:

$$x_{n+1} = A_{n+1,n} \cdot x_n + u \cdot w_{n+1}$$
$$y_n = C \cdot x_n + v_n$$

Where C is a matrix of the system, w_{n+1} is the noise indoor and v_n is the noise outdoor.

The first of these equations is known as process equation and the second one as measurement equation.

The first equation shows the relation among previous states and futures states, while the second one gives us the correspondence between the internal state of the system and how it can be observed.

These equations are useful for most of the linear estimation methods, like the Kalman filter described above.

The state-space representation requires two additional assumptions: the initial state vector has a known mean and variance and, besides, the perturbations w_{n+1} and v_n aren't correlated among them or with the initial state.

The system formed by the equations is linear, and for each moment, n and y_n can be expressed as a linear combination of the present and past values of w_{n+1} and v_n and the initial state vector.

The state-space representation using Kalman filter is calculated through a recursive procedure. The optimum estimator of the state vector for each moment n is based on the information available until this moment. This estimator is optimum because it minimizes the mean square error.

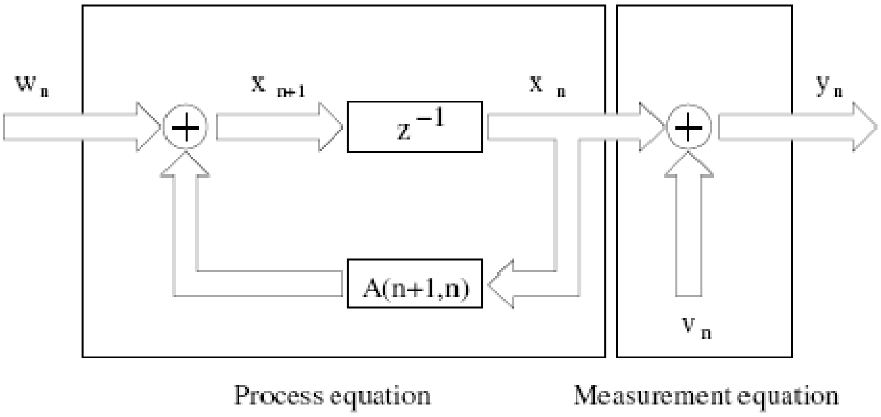


Figure 3.3: The signal flow of the Kalman filter. The function z^{-1} refers to a unit-delay operator. (Tom Bäckström, Speech Mathematics, 2007)

Among the advantages of the state-space model:

- Complete control over the dynamic of the model
- No loss of generality because the variables can be defined with past or future samples.
- It separates the sources of error and this allows that the stochastic part of the model has different effects.

The interpretation of w_n and v_n is important. The last one is a measure error, while w_n is described as the signal and defines the stochastic behavior of the model part that changes over the time.

3.2.5 The Kalman filter: advantages and disadvantages

3.2.5.1 Advantages

It avoids the influence of possible structural changes on the result. The recursive estimation starts from an initial sample and updates the estimations by adding a new observation until the end of the data. This implies that the most recent coefficients estimation is affected by the distant history; in presence of structural changes the data series can be cut. This cut can be corrected through the sequential estimations but with a biggest standard error. Like this, the Kalman filter, like other recursive methods, uses all the series history but with one advantage: it tries to estimate a stochastic path of the coefficients instead of a deterministic one. In this way it solves the possible estimation cut when structural changes happen.

The Kalman filter uses the least square method to recursively generate a state estimator on k moment, which is unbiased minimum and variance linear. This filter is in equal terms with Gauss-Markov theorem and this gives to Kalman filter its enormous power to solve a wide range of problems on statistic inference.

The filter is distinguished by its skill to predict the state of a model in the past, present and future, although the exact nature of the modeled system is unknown.
The dynamic modeling of a system is one of the key features which distinguish the Kalman method.

3.2.5.2 Disadvantages

Among the filter disadvantages we can find that it is necessary to know the initial conditions of the mean and variance state vector to start the recursive algorithm. There is no general consent over the way of determinate the initial conditions.

The Kalman filter development, as it is found on the original document, is supposed a wide knowledge about probability theory, specifically with the Gaussian condition for the random variables, which can be a limit for its research and application.

When it is developed for autoregressive models, the results are conditioned to the past information of the variable under study. In this sense the prognostic of the series over the time represents the inertia that the system actually has and they are efficient just for short time term.

Design and development

4. The project

The main idea in this project is to recover the clean speech signal from a sample corrupted with background noise through a telephone conversation. To achieve our goal we are going to estimate the speech spectrum using LP method (as Gannot proposed) and using SWLP, an improved algorithm developed at Helsinki University of Technology. As we mentioned above, with this method, choosing correctly the parameters we can obtain an improved robustness against the background noise. After this, the idea is to use the Kalman filtering as a tool to estimate the future clean samples from the first one in an iterative way. Finally we are going to compare both: the Gannot's way and our way.

4.1 The material

The only materials we need to develop this research are the recordings corrupted with noise. We are going to use a noisy speech corpus (NOIZEUS), which was developed to facilitate the researches over speech enhancement algorithms. The noisy database contains 30 IEEE sentences, produced by three male and three female speakers, and corrupted by eight different real environment noises at different SNRs. The noise was taken from the AURORA database and includes suburban train noise, babble, car, exhibition hall, restaurant, street, airport and train station noise.
The results showed in the next chapter are obtained with suburban train noise.

The IEEE database contains phonetically-balanced sentences with relatively low word-context predictability. The sentences selected from this database for NOIZEUS include all phonemes in the American English language.
The sentences were originally sampled at 25 kHz and downsampled to 8 kHz.

To simulate the frequency characteristics of telephone handsets, the speech and noise signals were filtered by the modified Intermediate Reference System filters used in ITU-T P.862.

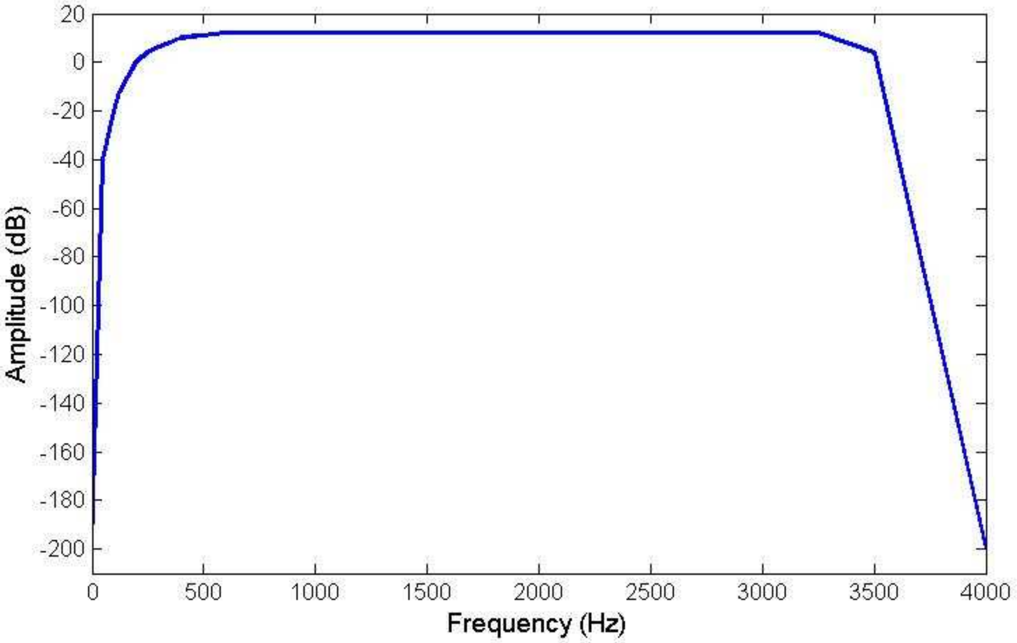


Figure 4.1: Frequency response of IRS filter

Noise is artificially added to the clean speech signal. The IRS filter is applied to the clean and noise signals independently. The speech level of the filtered clean signal is determinate. Afterwards, a noise segment is randomly cut with the same length of the speech signal from the noise recordings, and then it is scaled to reach the SNR level and finally added to the filtered clean speech signal.
In this database we can find signals at SNRs of 0 dB, 5 dB, 10 dB and 15 dB.

4.2 The program's code: step by step

All the code has been developed with Matlab, which is a high-performance language for technical computing. It integrates computation, visualization, and programming in an easy-to-use environment where problems and solutions are expressed in familiar mathematical notation.

4.2.1 First option: assuming white noise

Once we have loaded with the files the original clean signal and the signal corrupted with noise, we look for the amount of noise added. As we know from the recording's database, the noise is after added; then, we can extract it easily by calculating the difference between the clean speech signal and the noisy signal.

In the first option, as it will be described in the next chapter, we consider white noise; the coefficients of the noise added signal are calculated through LPC method and we use this estimation for all the process, assuming they don't change.

We are working with voice signals which are not stationary long-term, so they change quickly. To study them is better to take frames with lower time duration, in our case 30 ms. In this way we achieve smooth transitions between sounds and a quasi-stationary signal's behavior.
In our program, we get smaller signal from the original windowing it with an algorithm.

This algorithm uses the Hanning window,

$$w[n]=\begin{cases} 0.5-0.5\cdot\cos(2\cdot\pi\cdot n/M) & ,0 \le n \le M \\ 0 & ,\ \text{the rest} \end{cases}$$

where M is the length of our window, in this case, 240 samples.

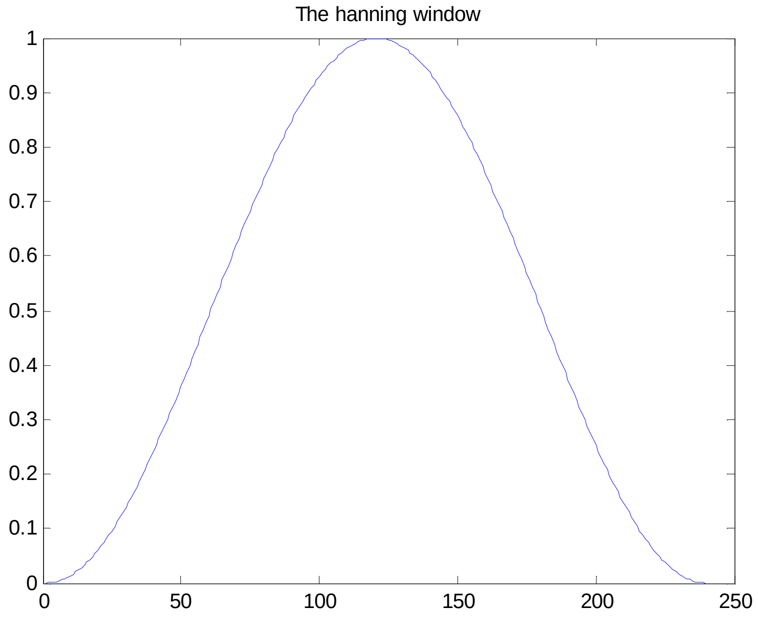


Figure 4.2: The Hanning Window

On time domain, we can observe the window takes lower values toward the edges, until zero. The samples in this area are much diminished and we lose information. For that reason, in our algorithm we are going to overlap the contiguous windows. In this way the quality of the results will be better.

We window therefore the original signal with added noise as the signal which just contains the noise, and we save them in two matrix where

each row contains the values of each window. Our window length is 240 samples.
Using a loop to cover all the windows of the signal, we calculate for each one, the coefficients of the original noisy signal with the LPC or SWLP method. It depends on which kind of results we want to get or compare.

At this very moment is when we use our Kalman filter, programmed as explained in the previous chapter.
Afterwards, all the necessary updates are done because, as we have studied, the past samples have an influence on the future ones.

Finally, the SNR ratio of the final signal is calculated after filtering. This will be the graph we will use to compare the different methods and options.

All this process is done some times, in our project eight times, because of the recursive characteristic method. The samples obtained are closer to the original ones through the time, because the past samples influence on future samples.
We did some tries increasing the number of the iterations, but the profit was not enough compared with the computational weight.

4.2.2 Second option: assuming colored noise

As we observed the results weren't those which we expected, we decided to get closer to the real life noise by assuming colored noise. The only difference is that in this case, we model noise characteristics by an AR-model in a similar way as we did with the speech before.

The only difference in the code with this change is that the coefficients of the noise added signal change with the time. For that reason, inside the loop, each time we calculate the coefficients of the original signal with noise, we have to calculate the coefficients of the noise signal too; but in this case we will use just the LPC method for all the tries.

Due to this change, we have to calculate the noise added to the signal, each time we repeat the process (in our project eight times).

4.3 The evaluation of the algorithms

To evaluate the results of the research, we are going to use three types of graphs:

- The representation of the voice signals amplitude as a function of time. This method is good to observe the voice signal shape, if it has so much noise from the original one or not and if it is much distorted; but with this kind of representations it is impossible to appreciate which method is better than the other.

- The spectrograms of the signals using the wavesurfer 1.8.5 tool. These are the representation of the energy signal as a function of time and frequency. On these figures it is possible to observe how Kalman works, how it is better on the speech frames than on the silent frames. Here again it is quite difficult to observe the difference between the two methods we want to compare, for that we are going to use the third kind of graph.
- The representation of the SNR method. On these figures we can see the SNR level (dB) of the signal as a function of values proportional to the time. As we can separate the speech signal estimated from the noise signal estimated, it is possible to calculate the signal noise ratio of each window through the calculation of the energies of each one and after translate it into dB's. This graph is the best to observe in an objective way, the differences between both methods.

Results and tests

5. Results

The signal that we use for this project is corrupted with the noise of a train, and the sentence said on it is:

"The birch canoe slid on the smooth planks"

With these kinds of sentences that we use to study the speech, the meaning of them is not as important as its balancing, phonetically speaking. In concrete, this sentence is pronounced by a man and sampled at 25 khZ.

For all the next experiments we are going to use the parameter h, the length to calculate the coefficient of the noise signal, with a value of 10.

In this figure we have the originals signals: the first one is the signal without noise, just the man's voice, and the rest are signals with different levels of background noise, SNR= 15dB, 10dB, 0dB. We are going to study along this project how our research works through all of them.

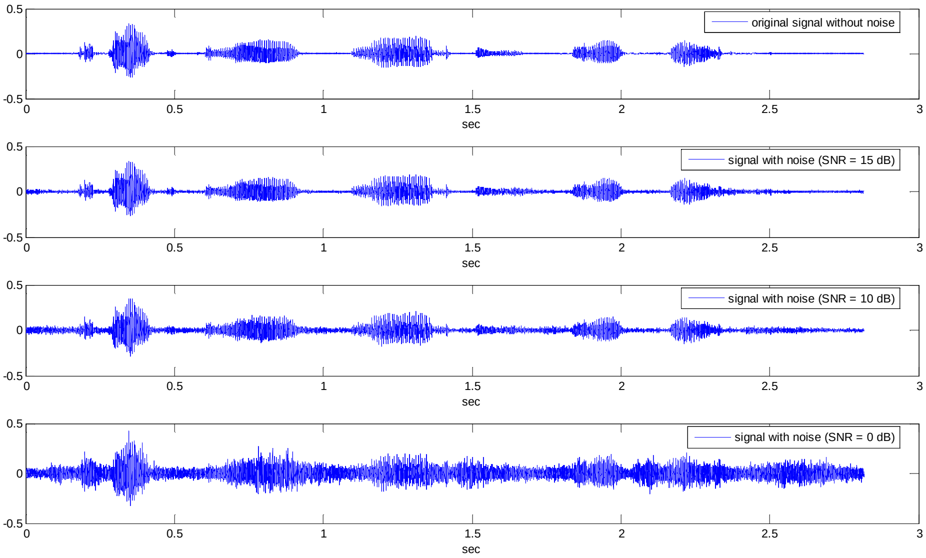


Figure 5.1: Original signal and signals with different levels of noise: SNR= 15, 10, 0dB before filtering

To appreciate better the different levels of noise we are going to include the spectrogram of these signals.

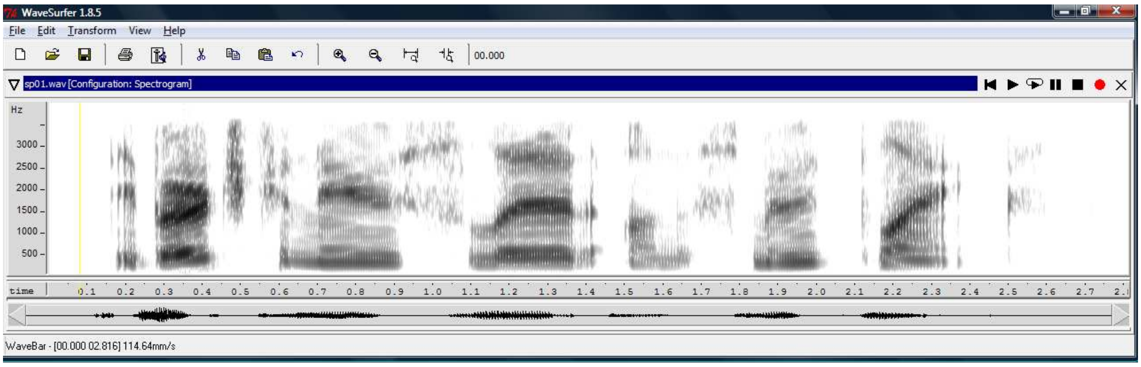


Figure 5.2: Original signal spectrogram

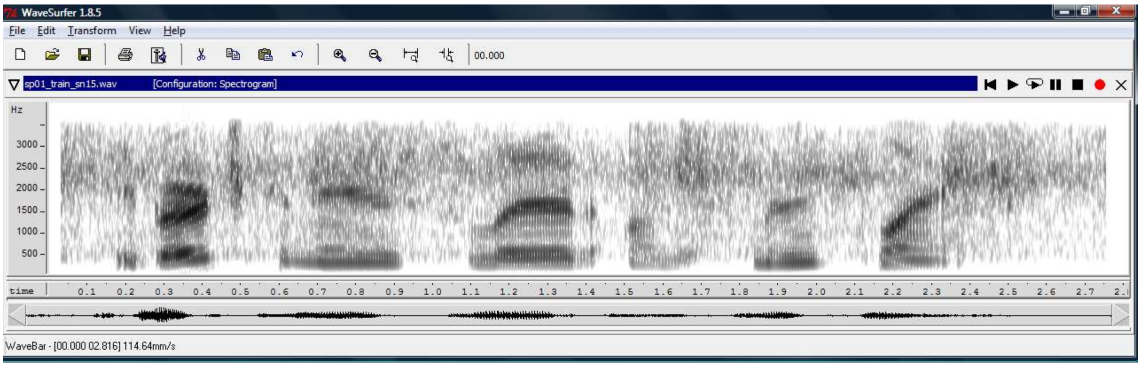


Figure 5.3: Spectrogram of the signal with SNR= 15dB before filtering

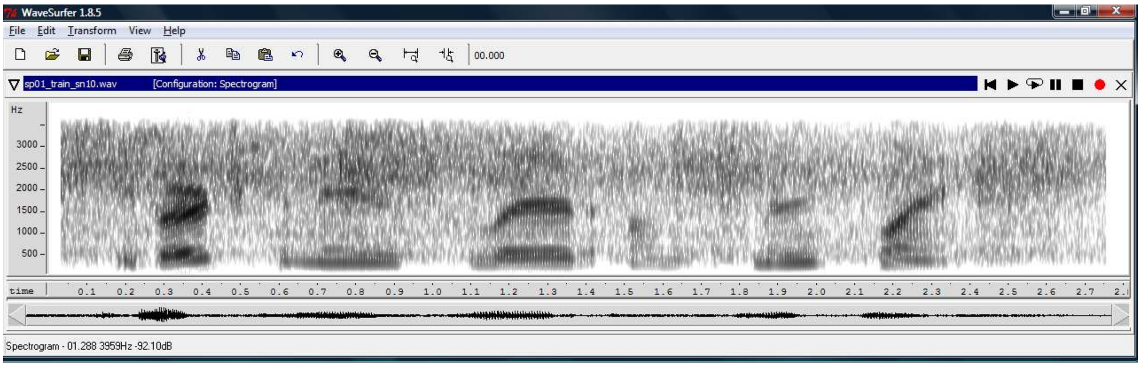


Figure 5.4: Spectrogram of the signal with SNR= 10dB before filtering

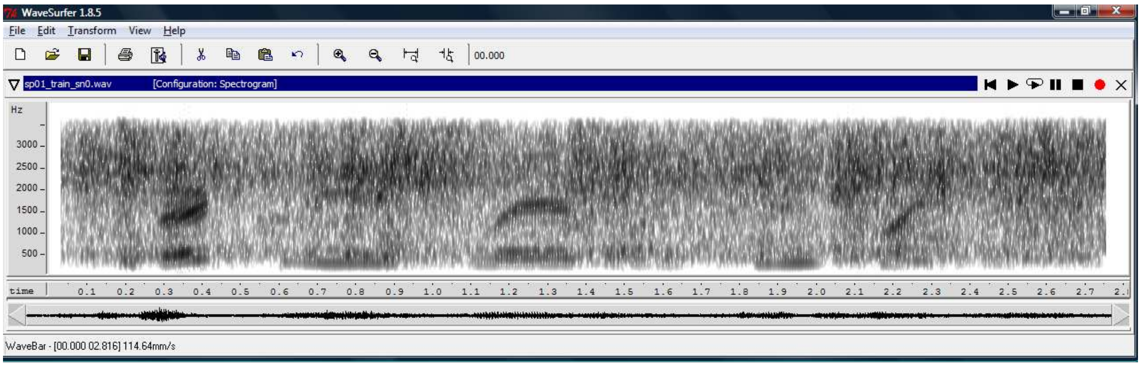


Figure 5.5: Spectrogram of the signal with SNR= 0dB before filtering

With these signals we can appreciate perfectly the different levels of noise. In the first one we can see the difference between the area with speech (dark color) and the area with silent (white color). In the rest of them we can observe how we have more noise in each one; the areas where we should have silence are becoming darker and darker as we add more noise.

As the SNR level is decreasing, it is harder to see where the voice is. With SNR = 10dB we can understand something but with SNR = 0dB, it is almost impossible to distinguish the speech from the silent frames because the noise level is very high.

5.1 First approximation: White noise

In our first research, we are going to assume white noise. This means we have the same noise power for all frequencies. We are going to calculate the coefficients of the noise added to the original signal, at the beginning of our algorithm and then use them for all the development; that is, the LPC coefficients of the noise are not going to change as the differences between the noise signal and the final signal change.

5.1.1 Prediction of the signal coefficients with LPC

Our first step is to use Gannot's method; this is, linear predicting coding (LPC) to predict the coefficients of the speech with noise signal. The results after Kalman filtering are the following.

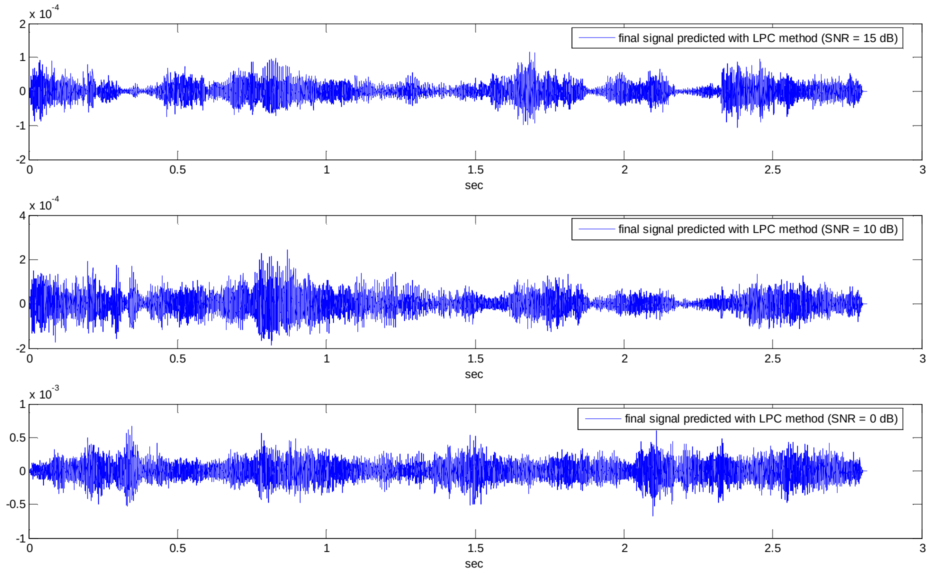


Figure 5.6: Final signals, using LPC as prediction method, from original signal with different noise levels (SNR = 15, 10, 0 dB) and noise assumed as white

If we compare the results of the figure 5.6 with the figure 5.1, could resemble that the improvement doesn't exist, but this is because the representation of the signal is not a good method to see the improvement.
The only visible difference is the amplitude in the figure 5.6 is lower than in the figure 5.1. This is because Kalman filtering has two sigma parameters, one of them is to calculate the part of LPC and the other one is for the noise part. These parameters represent the energy that we assume that the original signal with noise and the noise signal have. If we increase the parameter corresponding to the noise signal, then Kalman filtering will try to put more amount of noise into the noise signal. This means that more noise will be removed, which is the goal of our system.
However, on the other hand, Kalman filtering will remove some parts of the speech which resemble noise, corrupting the final speech signal.
In this way, changing the values of sigma, we can control how much noise is removed and how much speech is corrupted.

As we mentioned above, sigma parameters represent the amount of energy that each part of the signal has. If we assume that the original signal with noise has low energy, after the Kalman filter the final signal will have low energy too. This is the reason of the low amplitudes of the figure 5.6. In our physical system this is translated into a low volume of the signal and can be solved multiplying the final signal by a constant (1000 for example) increasing in this way the volume.

We can observe the changes easily between the figures 5.1 and 5.6 with the spectrograms.

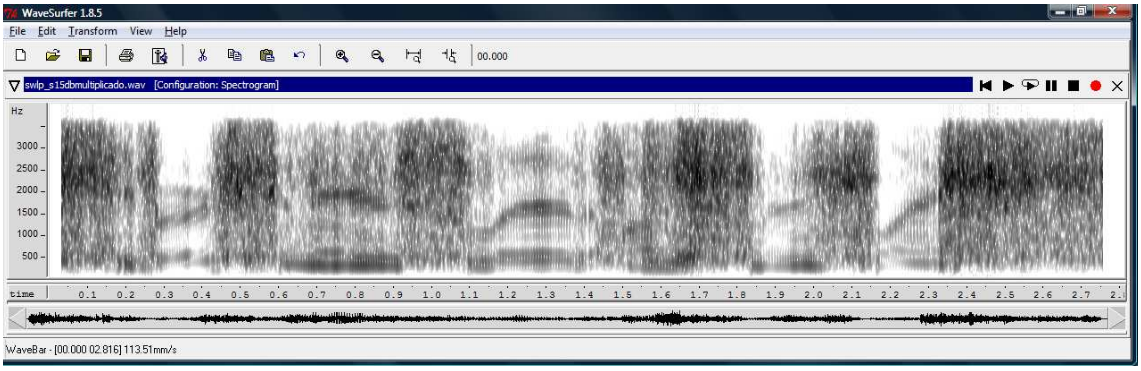


Figure 5.7: Spectrogram of the signal with SNR= 15dB after Kalman filtering and using the LPC method to predict the coefficients (white noise assumed)

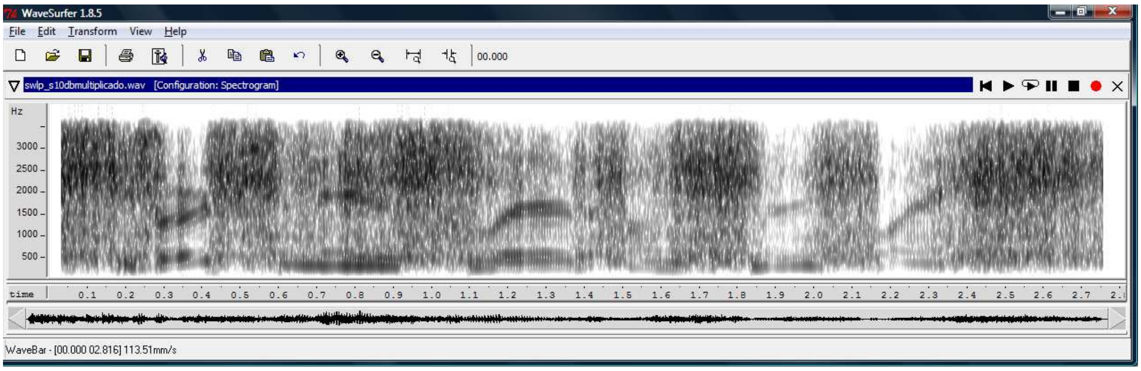


Figure 5.8: Spectrogram of the signal with SNR= 10dB after Kalman filtering and using the LPC method to predict the coefficients (white noise assumed)

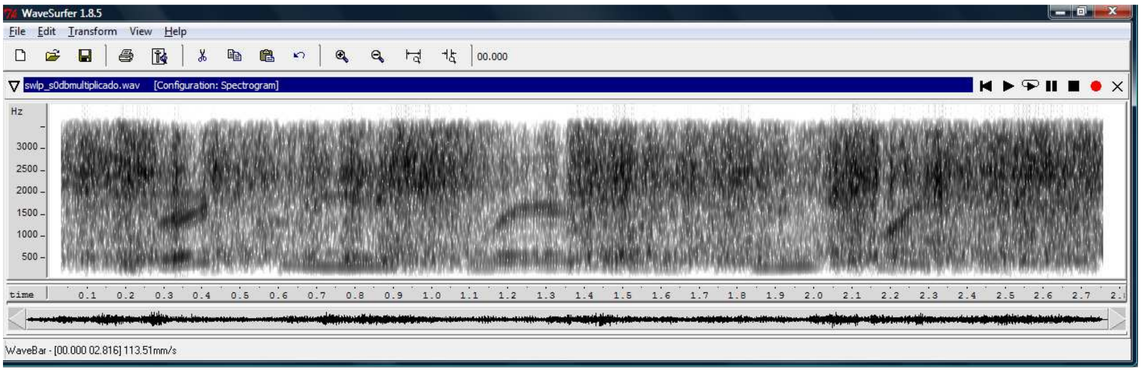


Figure 5.9: Spectrogram of the signal with SNR= 0dB after Kalman filtering and using the LPC method to predict the coefficients (white noise assumed)

In these figures the signals have been multiplied by 1000 as we said before.

Comparing these figures with the figures 5.2-5.5, we get less noise in the frames where we have speech. This is because we estimate the LPC model for each frame, and the model captures speech and noise. Besides, we assume that we have the model of the noise. Kalman then, tries to separate two different signals, the speech with noise signal and the signal with just noise. Like we have separated one part of noise, then the speech signal with noise has less noise, which is our goal.

If we do the same process for the part where we have only noise, there is not any difference between our noise model and the model of this frame. In this case, Kalman filter will be unable to remove the noise.
For that reason, in the parts where we have only noise, we will have the same amount of it after Kalman filtering.

As we can see through these results, after Kalman filtering we have less background noise and we will understand better the conversation.

5.1.2 Prediction of the signal coefficients with SWLP

Now, we are going to go one step further, and instead of using the LPC method to predict the signal's coefficients, we will introduce the SWLP algorithm in our system.
We still consider only white noise.

We continue working with the same signals as input (the original one, and the signals corrupted with noise with the same levels of SNR). For this reason, it is not necessary to draw them again (they are from Figure 5.1 to Figure 5.5).

The final signals after use the Kalman filter are these:

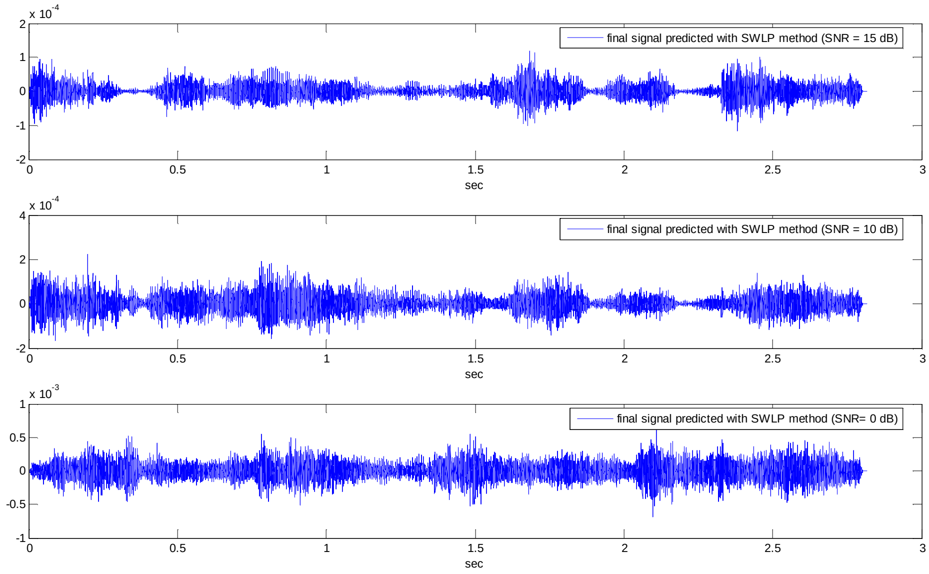


Figure 5.10: Final signals, using SWLP as prediction method, from original signal with different noise levels (SNR = 15, 10, 0 dB) and noise assumed as white

At a glance, we can't judge the results of this method; we are going to use the spectrograms as before to see them clearly.

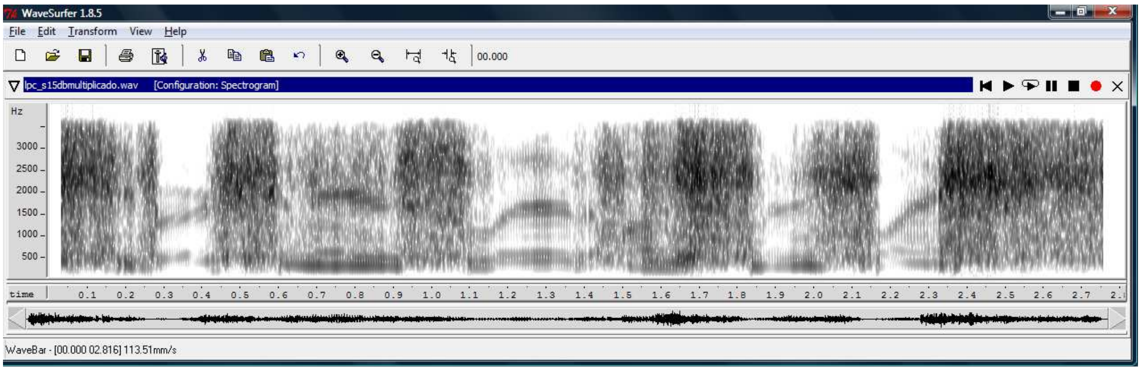

WaveSurfer 1.8.5
File Edit Transform View Help
00.000
lpc_s15dbmultiplicado.wav [Configuration: Spectrogram]
Hz
3000
2500
2000
1500
1000
500
time 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0 1.1 1.2 1.3 1.4 1.5 1.6 1.7 1.8 1.9 2.0 2.1 2.2 2.3 2.4 2.5 2.6 2.7
WaveBar - [00.000 02.816] 113.51mm/s

Figure 5.11: Spectrogram of the signal with SNR= 15dB after Kalman filtering and using the SWLP method to predict the coefficients (white noise assumed)

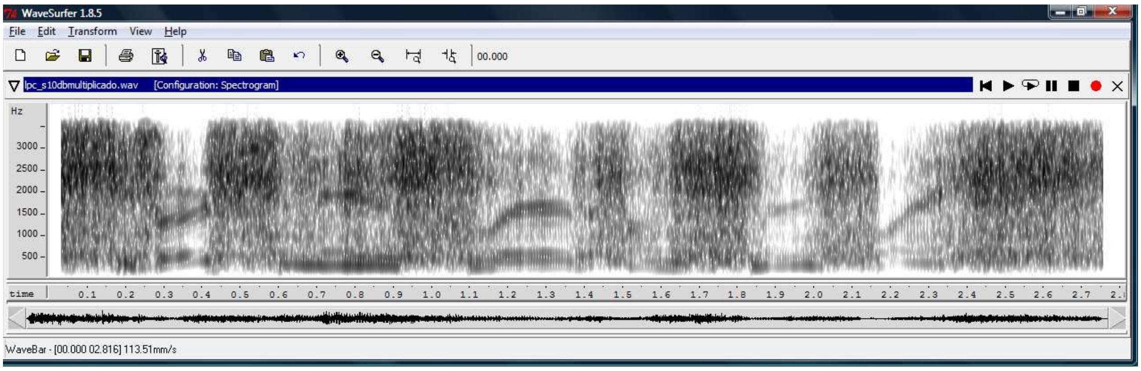

WaveSurfer 1.8.5
File Edit Transform View Help
00.000
lpc_s10dbmultiplicado.wav [Configuration: Spectrogram]
Hz
3000
2500
2000
1500
1000
500
time 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0 1.1 1.2 1.3 1.4 1.5 1.6 1.7 1.8 1.9 2.0 2.1 2.2 2.3 2.4 2.5 2.6 2.7
WaveBar - [00.000 02.816] 113.51mm/s

Figure 5.12: Spectrogram of the signal with SNR= 10dB after Kalman filtering and using the SWLP method to predict the coefficients (white noise assumed)

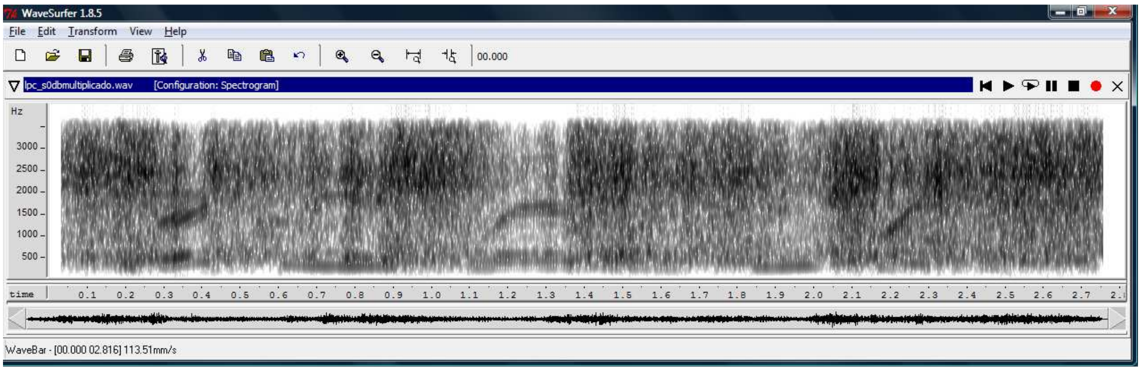

WaveSurfer 1.8.5
File Edit Transform View Help
00.000
lpc_s0dbmultiplicado.wav [Configuration: Spectrogram]
Hz
3000
2500
2000
1500
1000
500
time 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0 1.1 1.2 1.3 1.4 1.5 1.6 1.7 1.8 1.9 2.0 2.1 2.2 2.3 2.4 2.5 2.6 2.7
WaveBar - [00.000 02.816] 113.51mm/s

Figure 5.13: Spectrogram of the signal with SNR= 0dB after Kalman filtering and using the SWLP method to predict the coefficients (white noise assumed)

As before, we can observe that the final signal has less noise than the original one in the frames where we have human voice, but we can't measure how much better the filter is, and if this method is better or worse than LPC method to predict the coefficients of the original signal.

In this second option, using SWLP to estimate the coefficients, we can see the improvement as we saw using LPC method. The final signal has less noise, and the conversation would be better.

Now we are going to see the differences of both methods: LPC and SWLP.

5.1.3 Comparison between both methods of coefficients prediction: LPC and SWLP

Here we arrive to the important stage of our research; we are going to compare the results of LPC as prediction method, as Gannot proposed, with the results of using SWLP as prediction method, as we propose in this project.

Watching the previous graphs, we can't notice so much the difference of both ways, so we are going to use the SNR (signal to noise ratio) method.
This method is very used to compare which speech signal is better. The ratio is the margin between the power of the transmitted signal, and the power of the noise. This measure is in dB. For this specific signals and results that we are getting, the SNR method is the best one to observe the difference.

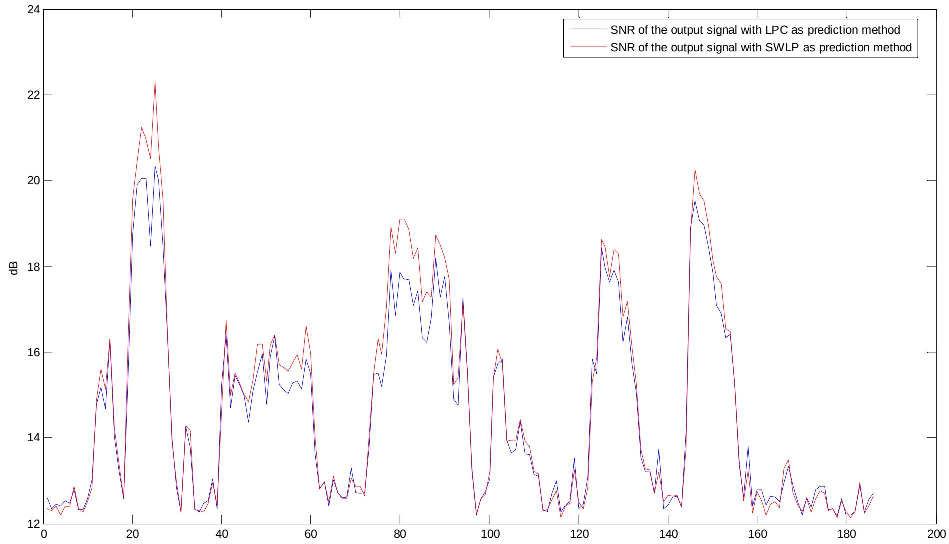


Figure 5.14: Comparison of the SNR's output signal using LPC or SWLP as prediction method when the input signal has SNR=15 dB (white noise assumed)

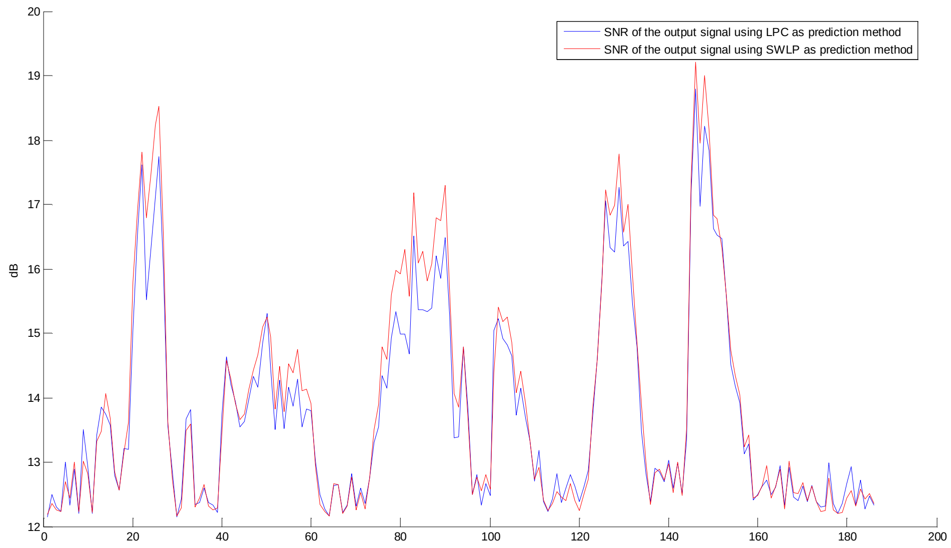


Figure 5.15: Comparison of the SNR's output signal using LPC or SWLP as prediction method when the input signal has SNR=10 dB (white noise assumed)

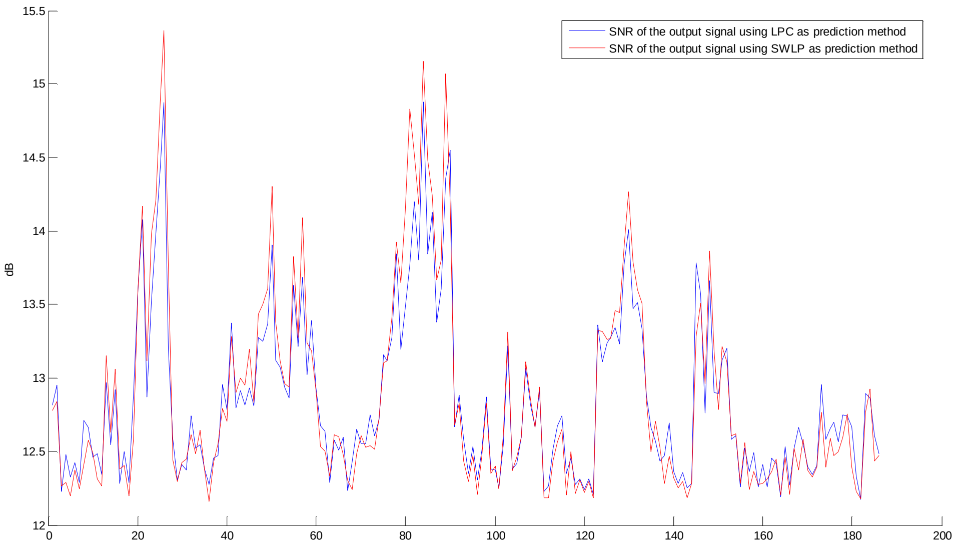


Figure 5.16: Comparison of the SNR's output signal using LPC or SWLP as prediction method when the input signal has SNR=0 dB (white noise assumed)

On these graphs we can observe how the SNR is higher when SWLP is as the method to predict the coefficients.

There are certain moments where the SNR when we use the SWLP method is a little lower than when the LPC method is used. However, this happens on the silent moments of the sentence, where there is not speech. As we showed before, it is on silent frames where the Kalman filtering is not optimum. Due to this, the combination of Kalman filter with the prediction of coefficients through SWLP method gives us the best results.

Anyway, we conclude that with this situation the results are very close; we can't say we have an improvement but neither a worsening because the differences are too small; they are not perceptible to the hearing.

After these experiments we decided to see what happened when the noise wasn't white, because in the real life noise hasn't the same power for all frequencies and the approximation that we did before was so coarse.

5.2 Second approximation: colored noise

In this section the noise is considered as not white and we have modeled noise characteristics by an AR-model in the same way that we did with the speech signal. Instead of calculating the LPC coefficients of the noise signal at the beginning, we will calculate the coefficients each time we update the output signal, this is because the noise is not always the same. We use the LPC method for noise coefficients because the SWLP method is optimized for speech.

The signals that we used as input are the same than in the previous section (Figures 5.1-5.5); the experiments are done with SNR = 15, 10, 0dB at the input.

As before, first we are going to calculate the coefficients of the speech with noise signal with LPC method, after with SWLP method and for ending we will compare both methods.

5.2.1 Prediction of the signal coefficients with LPC

The results after applying Kalman filtering are drawn in the figure 5.17. Here we can observe the output signals for the different values of SNR at the input.

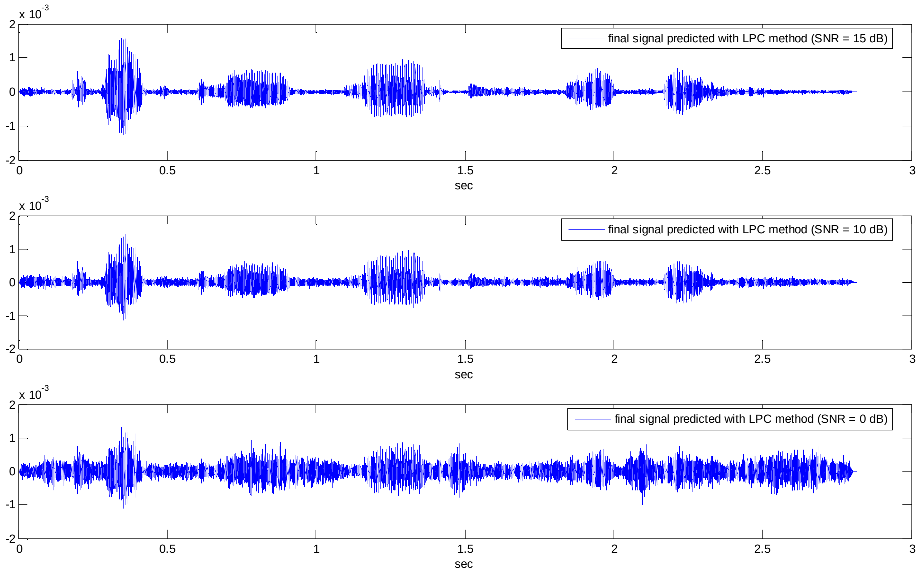
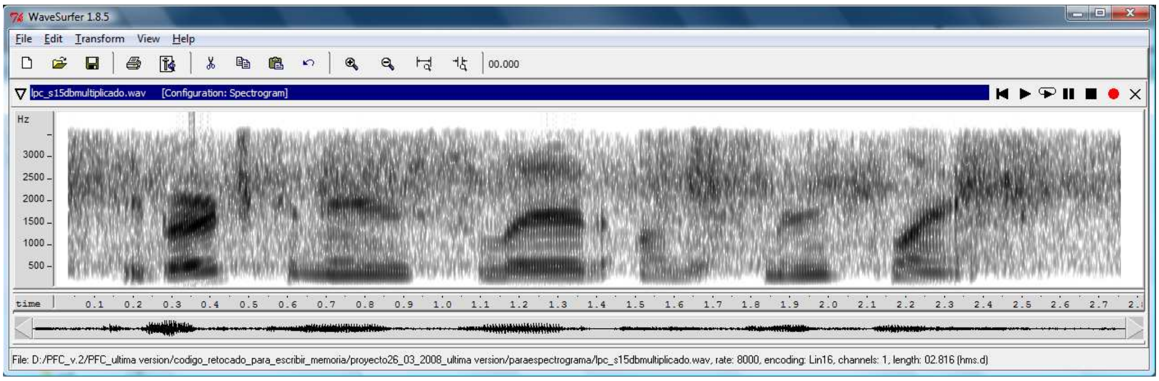


Figure 5.17: Final signals, using LPC as prediction method, from original signal with different noise levels (SNR = 15, 10, 0 dB) and noise assumed as colored

In this figure we can notice, like in the figure 5.6, that the amplitude of the output signals is much lower than in the input. This is because of, as we explained more detailed before, the effect of the sigma parameters. As we suppose low energy for the input signal, where we have speech and noise, then the output signal will have low energy too because of Kalman filtering, and this is translated into lower amplitude.

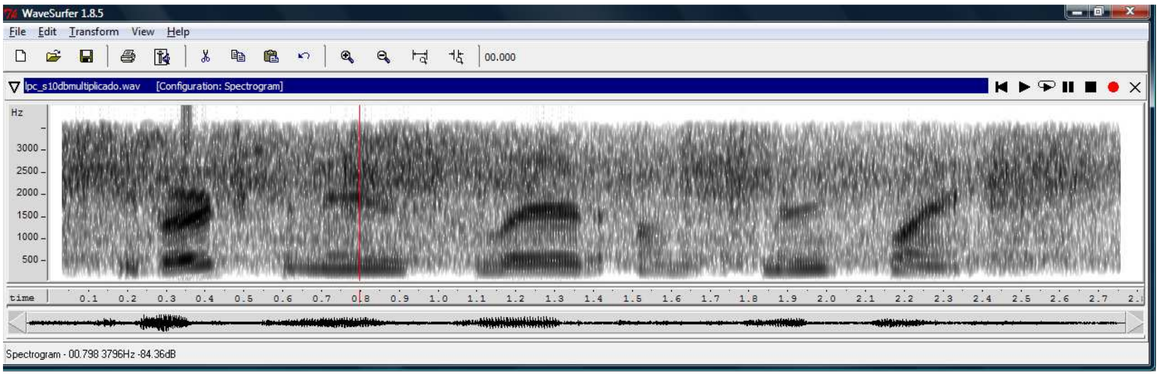
If we continue comparing the figures 5.6 and 5.17, it is very evident that in the second one we have less noise, and the final signal, after Kalman filtering, is less distorted. This is because the noise model now is not white, then, we don't have noise in all the frequencies, and those speech parts which are on frequencies without noise are not corrupted.



lpc_s15dbmultiplicado.wav [Configuration: Spectrogram]

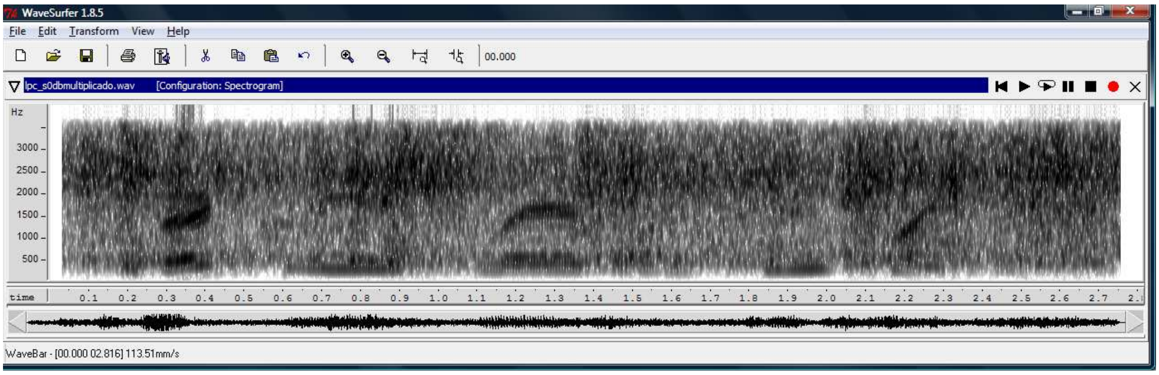
File: D:/PFC_v.2/PFC_ultima version/codigo_retocado_para_escribir_memoria/proyecto26_03_2008_ultima version/paraespectrograma/lpc_s15dbmultiplicado.wav, rate: 8000, encoding: Lin16, channels: 1, length: 02.816 (hms.d)

Figure 5.18: Spectrogram of the signal with SNR= 15dB after Kalman filtering and using the LPC method to predict the coefficients (colored noise assumed)


WaveSurfer 1.8.5
File Edit Transform View Help
00.000
lpc_s10dbmultiplicado.wav [Configuration: Spectrogram]
Hz
3000
2500
2000
1500
1000
500
time 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0 1.1 1.2 1.3 1.4 1.5 1.6 1.7 1.8 1.9 2.0 2.1 2.2 2.3 2.4 2.5 2.6 2.7 2.

Spectrogram - 00.798 3796Hz -84.36dB

Figure 5.19: Spectrogram of the signal with SNR= 10dB after Kalman filtering and using the LPC method to predict the coefficients (colored noise assumed)


WaveSurfer 1.8.5
File Edit Transform View Help
00.000
lpc_s0dbmultiplicado.wav [Configuration: Spectrogram]
Hz
3000
2500
2000
1500
1000
500
time 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0 1.1 1.2 1.3 1.4 1.5 1.6 1.7 1.8 1.9 2.0 2.1 2.2 2.3 2.4 2.5 2.6 2.7 2.

WaveBar - [00.000 02.816] 113.51mm/s

Figure 5.20: Spectrogram of the signal with SNR= 0dB after Kalman filtering and using the LPC method to predict the coefficients (colored noise assumed)

From figure 5.18 to 5.20 we can observe the spectrograms of the final signals which coefficients have been predicted with LPC model and after filtered by Kalman method.

If we compare these figures with figures 5.7-5.9, it is noticeable how the first ones have the speech part clearer, but the figures 5.18-5.20 have, on the speech parts more difference between the speech level and the noise level. This means that, although we can listen some noise on speech parts, we will understand the speech better. Besides, the noise level is constant through all the signal which is better for the human system to understand the speech.

5.2.2 Prediction of the signal coefficients with SWLP

The results after applying Kalman filtering are drawn in the figure 5.21. Here we can observe the output signals for the different values of SNR at the input.

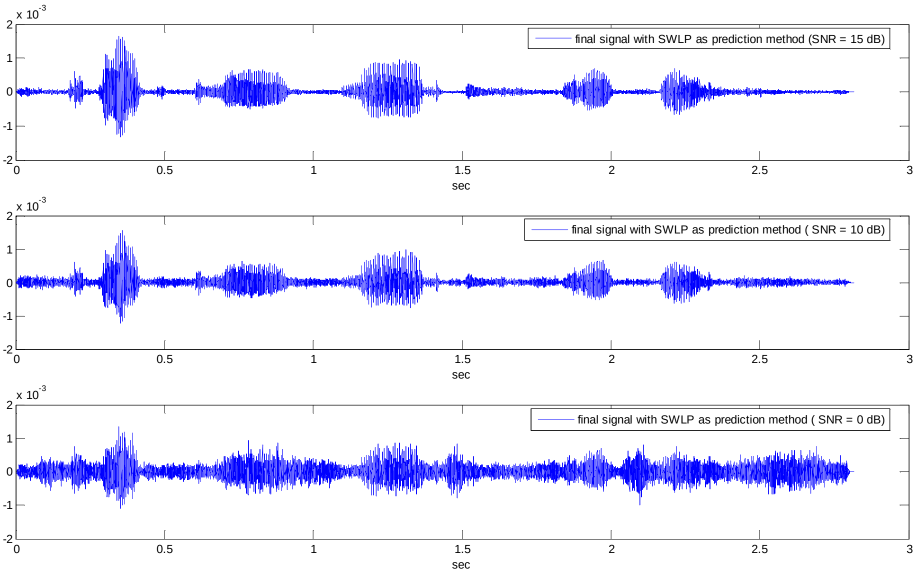


Figure 5.21: Final signals, using SWLP as prediction method, from original signal with different noise levels (SNR = 15, 10, 0 dB) and noise assumed as colored

Here, as it happened using the LPC method as prediction, the final signals are less corrupted by noise, and they are closer to the original ones, this is because we don't have noise in all frequencies.

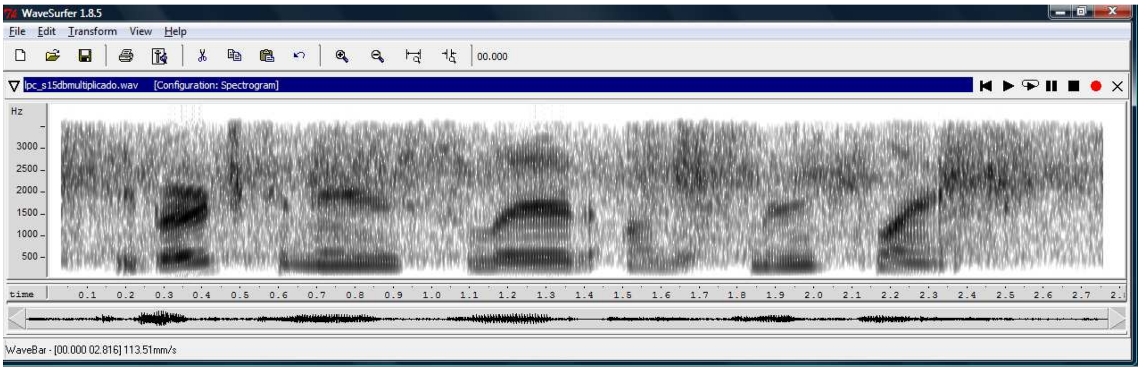


Figure 5.22: Spectrogram of the signal with SNR= 15dB after Kalman filtering and using the SWLP method to predict the coefficients (colored noise assumed)

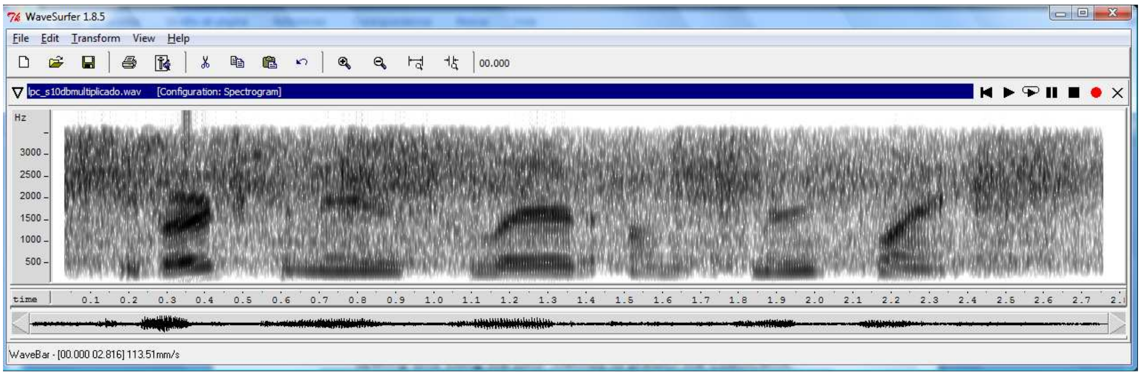


Figure 5.23: Spectrogram of the signal with SNR= 10dB after Kalman filtering and using the SWLP method to predict the coefficients (colored noise assumed)

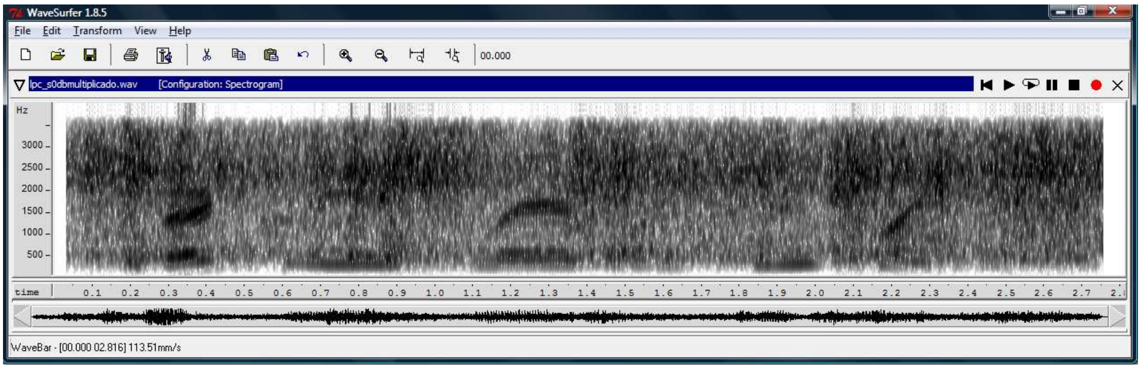


Figure 5.24: Spectrogram of the signal with SNR= 0dB after Kalman filtering and using the SWLP method to predict the coefficients (colored noise assumed)

Comparing the spectrograms we have the same result, figures from 5.11 to 5.13 have less noise in the speech parts and more noise in the silent parts than figures from 5.22 to 5.24. However, these later figures have the same level of noise in the speech parts and in the silent parts, which is better for the human ear to understand the speech, as it happens too for LPC method.
Here, the difference between the speech level and noise level is higher than with white noise, and the results will be better.

5.2.3 Comparison between both methods of coefficients prediction: LPC and SWLP

Again we are going to compare both coefficient prediction methods, LPC and SWLP, but with the noise modeled as colored.

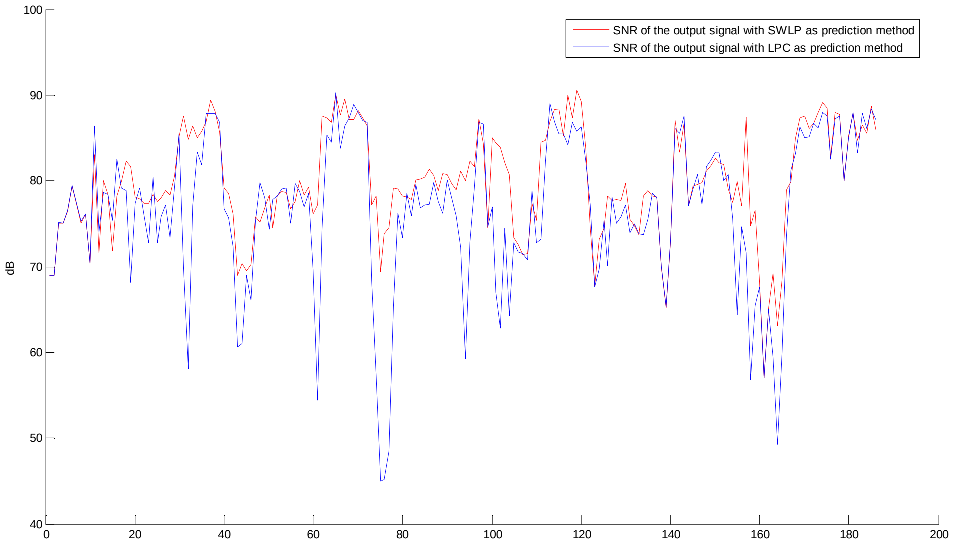


Figure 5.25: Comparison of the SNR's output signal using LPC or SWLP as prediction method when the input signal has SNR=15 dB (colored noise assumed)

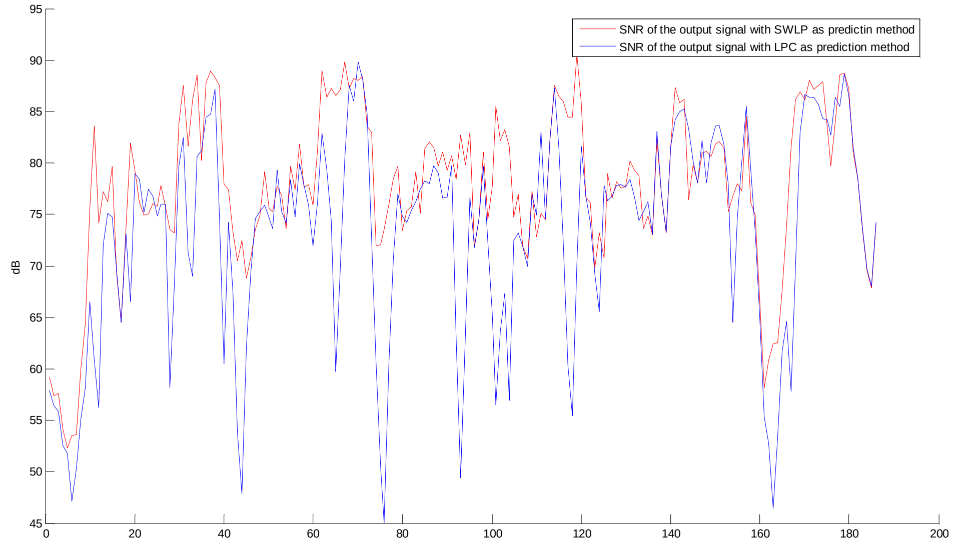


Figure 5.26: Comparison of the SNR's output signal using LPC or SWLP as prediction method when the input signal has SNR=10 dB (colored noise assumed)

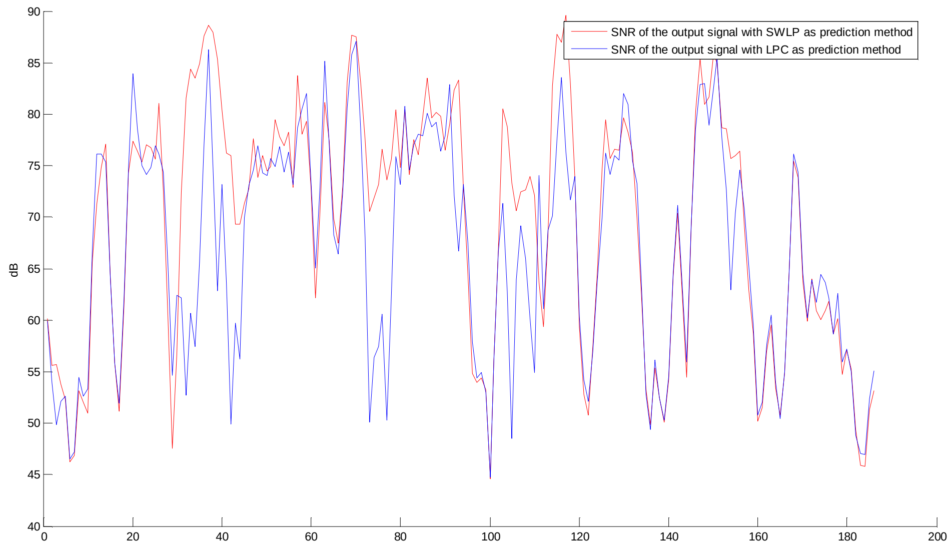


Figure 5.27: Comparison of the SNR's output signal using LPC or SWLP as prediction method when the input signal has SNR=0 dB (colored noise assumed)

In general, for both prediction methods, LPC and SWLP, we have more SNR in the entire signal.

The noise level along the signal is more constant but this could be good for the human hearing system, because the changes between the speech and the noise are smoother, and for that reason we can understand better the conversation.

We can see too how the SWLP prediction method has higher SNR than LPC. This is especially interesting on speech frames, although this difference is still very low for the human ear.

Conclusions and future work

In this project we have studied the effect of the Kalman filtering over telephonic conversations using two methods of coefficient prediction:

- The method proposed by Gannot: LPC
- The method developed and proposed by Helsinki University of Technology: SWLP

Our first step was to mount a system assuming a white noise model; considering the spectrograms and the figures with the SNR representation, we can observe how Kalman works better for the speech frames. It is easier to see with the SNR levels how this level increases for the speech and how it decreases for the noise frames.

The problem for this case is exactly that the SNR level changes abruptly from the speech to the noise frames, and for the human ear, which has kind of a memory it is very difficult to make this change fast enough to understand perfectly the conversation.

Taking into account the SNR representation for the noise modeled as white, we observed how the combination of Kalman filtering with SWLP has better results that with LPC method. This improvement is more noticeable on speech frames, where Kalman and SWLP method are both optimum.

On the second part, where we model the noise as colored, we could notice how the final signals, after Kalman filtering, where less corrupted by noise. This is because for Kalman is easier to eliminate noise, even in the frames without speech, when the noise model is better.

As a conclusion with the prediction methods, the SWLP is a bit better than LPC, but the difference, if we listen the recordings, is negligible for the human ear.

We have seen that on silent parts, the noise is still very high; for that it would be very useful for a real application to include a system to detect the speech and the silent frames. To do this we have voice activity detection (VAD) algorithms which detect the presence or absence of human speech. In this way we could make frames without speech in silent. The problems with these algorithms are the delay, sensitivity, accuracy and computational cost.

Conclusiones y trabajo futuro

En este proyecto hemos estudiado el efecto del filtrado de Kalman sobre conversaciones telefónicas utilizando dos métodos de predicción de coeficientes:

- El método propuesto por Gannot: LPC
- El método desarrollado y propuesto por Helsinki University of Technology: SWLP

Nuestro primer paso fue montar un sistema asumiendo un modelo de ruido blanco; considerando el espectrograma y las figuras en donde se representa el nivel de SNR, podemos observar cómo Kalman trabaja mejor para los marcos donde tenemos speech. Es más fácil ver con la representación del SNR cómo estos niveles aumentan para el speech y cómo disminuyen para los marcos con ruido.

El problema en este caso es que el nivel de SNR cambia abruptamente cuando pasamos de los marcos con speech a los marcos con ruido, y para el oído humano, que tiene memoria auditiva, le es muy difícil realizar este cambio de forma tan rápida como para que la conversación sea perfectamente entendible.

Teniendo en cuenta las representaciones de SNR para el modelado de ruido como blanco, observamos cómo la combinación del filtrado de Kalman con el método SWLP da mejores resultados que con el método LPC. Esta mejora se puede apreciar más en los marcos de speech, donde tanto Kalman como el método SWLP son óptimos.

En la segunda parte del documento, donde hemos modelado el ruido como coloreado, se puede apreciar cómo las señales finales, tras haber aplicado el filtrado de Kalman, están menos corrompidas por el ruido. Esto es debido a que para Kalman es más fácil en este caso eliminar el ruido, incluso en los marcos sin speech, donde el modelo de ruido es mejor.

Como conclusión de los métodos de predicción, el SWLP es un poco mejor que el LPC, pero la diferencia, si se escuchan las grabaciones, es imperceptible para el oído humano.

A través de todas las investigaciones y pruebas realizadas, hemos visto que en las partes donde no existe speech, tan sólo silencio, el ruido es todavía muy alto; por lo que sería muy útil para aplicaciones reales incluir un sistema para detectar los marcos de speech y de silencio. Para realizar esto se pueden utilizar algoritmos VAD, los cuales detectan la presencia o ausencia de habla humana. El problema con estos algoritmos es el retardo, la sensibilidad, la agudeza y el coste computacional.

References

Alan V. Oppenheim, "Signals and Systems", Prentice Hall, 1997.

Alan V. Oppenheim, "Tratamiento de Señales en tiempo Discreto", Prentice Hall, 2000.

Andrew Blake, "State – Space Models and the Kalman filter: Application, Formulation and Estimation", Bank of England, 2002.

Andrew C. Harvey, "Forecasting, structural time series models and the Kalman filter", Cambridge, 1989.

AURORA Project Database (Aurora 4a)

C. Ma, Y. Kamp, L. Willems, "Robust Signal selection for Linear Prediction Analysis of Voiced Speech", Speech Communication, 12 (1):69-81, 1982.

Carlo Magi, Jouni Pohjalainen, Tom Bäckström, Paavo Alku, "Stabilised Weighter Linear Prediction", in Proc. Of Interspeech, Antwerp, Belgium, August 27-31 2007.

Deller, J.R., J.H.L. Hansen, J.G. Proakis, "Discrete – Time Processing of Speech Signals", IEEE Press, 2000.

J. Hynninen, "GuineaPig- a generic subjective test system for multichannel audio", in Proc AES 106th Convention, Munich, Germany, May 8-11 1999, Vol. 106.

J. Makhoul, "Linear prediction: A tutorial review", IEEE Proceedings, Volume 63, No. 4, Pages 561-580, 1975.

J.D. Hamilton, "Time Series Analysis", Princeton University Press, 1994.

Jacob Benesty, Shoji Makino, Jindong Chen (Eds.), "Speech enhancement", Springer, 2005.

Javier Ortega, Handouts of the subject "Temas Avanzados en Procesamiento de señales"

L.R. Rabiner, R.W. Schafer, "Digital Processing of Speech Signals", Prentice Hall, 1978.

M. Wölfel, J. McDonough, "Minimum Variance Distortionless Response Spectral Estimation", IEEE Signal Processing Magazine, 22(5): 117-126, 2005.

Mendel, J.M., “Lessons in Estimation Theory for Signal Processing, Communications, and Control”, Upper Saddle River, NJ, Prentice Hall, 1995.

Monson H. Hayes, “Statistical digital signal processing and modeling”, Wiley, 1996.

S. Gannot, D. Buhrshtein, E. Wienstein, “Iterative and sequential Kalman filter-based speech enhancement algorithms”, IEEE Trans. Speech Audio Processing, Vol. 6, No. 4, Pages 373-385, July 1998.

Sadaoki Furui, “Digital speech processing, synthesis, and recognition”, Marcel Dekker, 2001.

T. F. Quatieri, “Discrete – time speech signal processing principles and practice”, Prentice Hall, 2002.

Tom Bäckström, “Speech Processing Mathematics”, 2007.

Welch, Grag abd GaryBishop, “An Introduction to the Kalman Filter”, TR 95-041, Department of Computer Science, University of North Carolina at Chapel Hill, 2002.

www.mathworks.com

Xuedong Huang, Alex Acero, Hsiao-Wuen Hon, “Spoken Language Processing”, Prentice Hall, 2001.

Y. Hu, P. Loizou, “Subjective comparison of speech enhancement algorithms” in Proc. IEEE Acoustics, Speech, and Signal Proc. ICASSP-2006, Touluse, France, May 2006, Vol. I, Pages 153-156.

Appendix

A. Kalman's Biography and achievements

Rudolf Kalman was born in Budapest (Hungary) on May 19, 1930, the son of an electrical engineer. Early wanted to follow his father's model and started a career on mathematics field. He emigrated to United States on 1943 and studied electrical engineering at Massachusetts Institute of Technology (MIT) in Cambridge, receiving his master's degree in 1954.
Later he got his doctorate in 1957 at Columbia's University, in New York City. There, he had the fortune to study with the professor John R. Ragazzini, head of electronic laboratory and a important man for his research about ultra-high frequency (UHF) techniques, analog computers and control systems.

During his years in MIT and Columbia, Kalman explored his interest in control theory to study how to get, using mathematics, a device controlled to change the data stream output in a desired output.
Later, he worked at the Research Institute for Advanced Studies in Baltimore (RIAS) as mathematical researcher and as associate director of research.

Through lectures and published papers, he helped to extend the knowledge about the modern control theory, which includes programming robotics and machines to answer the constant change of the conditions and keep self-control: an application of this theory could be an automatic pilot system in a flight without crew to avoid a crash.

All Kalman's research had an important impact in these fields:

- Research about fundamental systems concepts: controllability and observability.
- Development of theories in structural aspects of system engineer
- Unify the theory and the design on linear systems with respect to quadratic criteria.

Besides, he was one of the first in using digital computer like an important part of the design process as well as of the control system's implementations.

However, the most important work for Kalman was the development of Kalman filters. At the beginning of his research he found solutions to problems with discrete-time filters. Kalman based his study on filters carried out by Norbert Wiener, Kolmogorov, Bode, Shannon and Pugachev among others. On the basis of state-space techniques and some recursive algorithms, the Kalman filter revolutionized the estimation field.

One of the forces which financed Kalman's work was the U.S. Air Force. By the late 1950s and early 1960s, the aircrafts had been developed to the point that they needed advanced control mechanisms for the flights. The AFORS sponsored a lot of researches in this area, even those which Kalman and Bucy did in RIAS. As we had said, his work revolutionized the estimation field and had an important impact in the design and development of navigation systems. The Kalman filter was the biggest discovery in orientation technology.
The Kalman algorithm was used by NASA. Firstly, they used the filter to solve some problems about satellites orbits, and later, it was included for Ranger, Mariner and Apolo missions; and when the Apolo 11 module was landed to the Moon in July of 1969, was guided by the Kalman filter.

Due to all this, the Kalman filter is the most utilized product in the modern control theory, and is used in almost every control system with commercials and militaries purposes:

- Navigational and guidance systems.
- Radar tracking algorithms for anti-ballistic missile applications.
- Sonar ranging.
- Satellite orbit determination.

B. Noise

Noise is an unwanted signal that interferes with the communication, the measure or the information processing which transports a signal. Technically, noise is the result of the combination of more than one sound with just one frequency and has a spectrogram with continuous frequency, with irregular amplitude and wave length.

The success in the noise processing method depends on the skill to characterize and model the noisy process, and use its characteristics to separate it from the signal to restore.

There exist a lot of noise types, but in this projects the main are the white noise and the colored noise.

White noise is a random signal with a flat power spectral density. It contains equal power at any frequency within a bandwidth. By having power at all frequencies, the total power of the signal is infinite, which is impossible, for that, it is said that white noise is a theoretical construction. However it is possible over a defined frequency band.

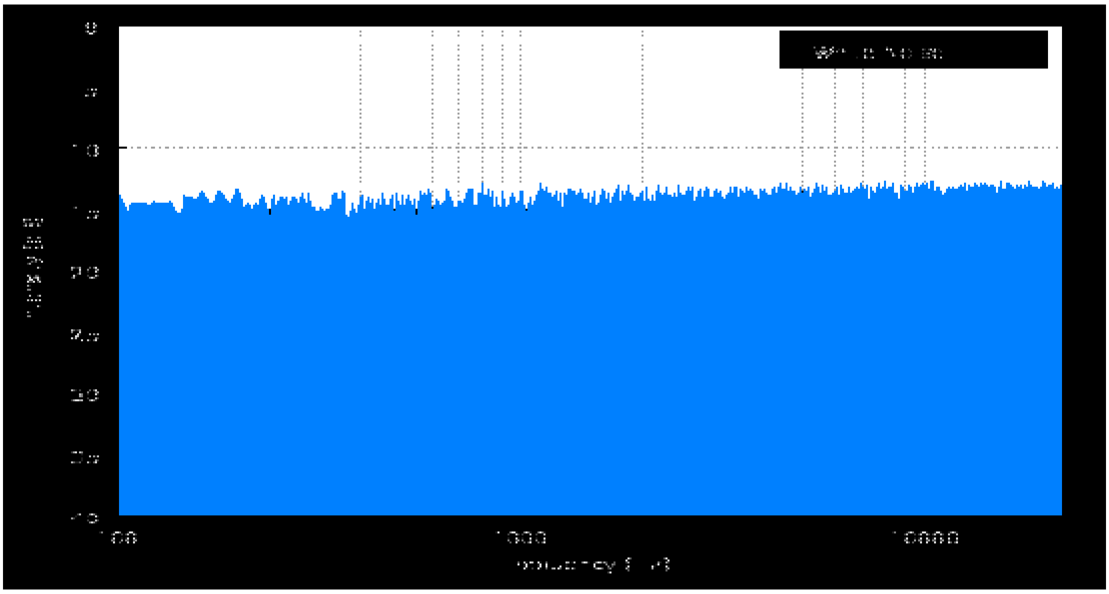


Figure B.1: White noise spectrum (Wikipedia.org)

Colored noise is a broadband noise with a spectrum different from the white noise spectrum. It continue being a random signal but with statistical characteristics and properties. Depending of the shape of its power spectral density, we have different colors for the noise.
The background noise has a spectrum which is not as the white one, with a predominance of low frequencies.

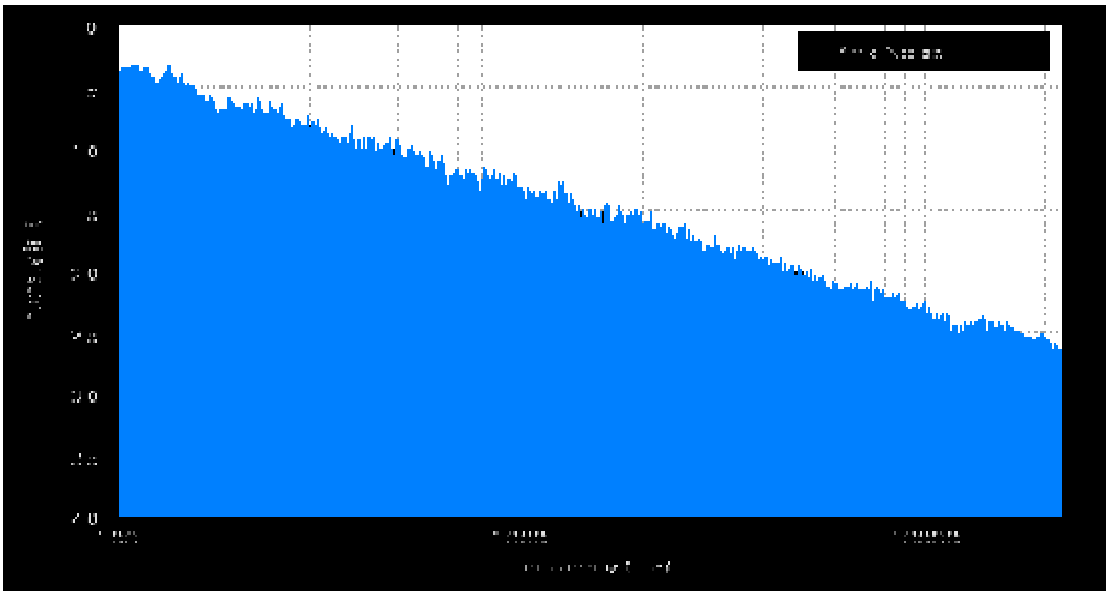


Figure B.2: Pink noise spectrum (Wikipedia.org)

BUDGET

1) Material Execution

- Purchase of personal computer (Software included)...... 2.000 €
- Rent a laser printer for 10 months... 100 €
- Office equipment .... 50 €
- Total material execution .... 2.150 €

2) Overheads

- 16 % on Material Execution .... 344 €

3) Industrial profit

- 6 % on Material Execution .... 129 €

4) Project Fees

- 640 hours 15 € / hour .... 9600 €

5) Consumables

- Printing Costs .... 100 €
- Binding.. 10 €

6) Subtotal budget

- Subtotal Budget .... 11860 €

7) I.V.A. applicable

- 16% Subtotal Budget .... 1897,6 €

8) Total Budget

- Total Budget .... 13757,6 €

Madrid, Octubre de 2008
El Ingeniero Jefe de Proyecto

Fdo.: Bárbara Valenciano Martínez
Ingeniero Superior de Telecomunicación

PRESUPUESTO

1) Ejecución Material

- Compra de ordenador personal (Software incluido)...... 2.000 €
- Alquiler de impresora laser durane 10 meses .... 100 €
- Material de oficina .... 50 €
- Total de ejecución material.. 2.150 €

2) Gastos generales

- 16 % sobre Ejecución Material .... 344 €

3) Beneficio industrial

- 6 % sobre Ejecución Material .... 129 €

4) Honorarios Proyecto

- 640 horas a 15 € / hora .... 9600 €

5) Material fungible

- Gastos de impresión .... 100 €
- Encuadernación.. 10 €

6) Subtotal del presupuesto

- Subtotal presupuesto... 11860 €

7) I.V.A. aplicable

- 16% Subtotal presupuesto... 1897,6 €

8) Total Presupuesto

- Total Presupuesto... 13757,6 €

Madrid, Noviembre de 2008
El Ingeniero Jefe de Proyecto

Fdo.: Bárbara Valenciano Martínez
Ingeniero Superior de Telecomunicación

PLIEGO DE CONDICIONES

Este documento contiene las condiciones legales que guiarán la realización, en este proyecto llamado "Speech enhancement using Kalman filtering". En lo que sigue, se supondrá que el proyecto ha sido encargado por una empresa cliente a una empresa consultora con la finalidad de realizar dicho sistema. Dicha empresa ha debido desarrollar una línea de investigación con objeto de elaborar el proyecto. Esta línea de investigación, junto con el posterior desarrollo de los programas está amparada por las condiciones particulares del siguiente pliego.

Supuesto que la utilización industrial de los métodos recogidos en el presente proyecto ha sido decidida por parte de la empresa cliente o de otras, la obra a realizar se regulará por las siguientes:

Condiciones generales

1. La modalidad de contratación será el concurso. La adjudicación se hará, por tanto, a la proposición más favorable sin atender exclusivamente al valor económico, dependiendo de las mayores garantías ofrecidas. La empresa que somete el proyecto a concurso se reserva el derecho a declararlo desierto.

2. El montaje y mecanización completa de los equipos que intervengan será realizado totalmente por la empresa licitadora.

3. En la oferta, se hará constar el precio total por el que se compromete a realizar la obra y el tanto por ciento de baja que supone este precio en relación con un importe límite si este se hubiera fijado.

4. La obra se realizará bajo la dirección técnica de un Ingeniero Superior de Telecomunicación, auxiliado por el número de Ingenieros Técnicos y Programadores que se estime preciso para el desarrollo de la misma.

5. Aparte del Ingeniero Director, el contratista tendrá derecho a contratar al resto del personal, pudiendo ceder esta prerrogativa a favor del Ingeniero Director, quien no estará obligado a aceptarla.

6. El contratista tiene derecho a sacar copias a su costa de los planos, pliego de condiciones y presupuestos. El Ingeniero autor del proyecto autorizará con su firma las copias solicitadas por el contratista después de confrontarlas.

7. Se abonará al contratista la obra que realmente ejecute con sujeción al proyecto que sirvió de base para la contratación, a las modificaciones autorizadas por la superioridad o a las órdenes que con arreglo a sus facultades le hayan comunicado por escrito al Ingeniero Director de obras siempre que dicha obra se haya ajustado a los preceptos de los pliegos de condiciones, con arreglo a los cuales, se harán las modificaciones y la valoración de las diversas unidades sin que el importe total pueda exceder de los presupuestos aprobados. Por consiguiente, el número de unidades que se consignan en el proyecto o en el presupuesto, no podrá servirle de fundamento para entablar reclamaciones de ninguna clase, salvo en los casos de rescisión.

8. Tanto en las certificaciones de obras como en la liquidación final, se abonarán los trabajos realizados por el contratista a los precios de ejecución material que figuran en el presupuesto para cada unidad de la obra.

9. Si excepcionalmente se hubiera ejecutado algún trabajo que no se ajustase a las condiciones de la contrata pero que sin embargo es admisible a juicio del Ingeniero Director de obras, se dará conocimiento a la Dirección, proponiendo a la vez la rebaja de precios que el Ingeniero estime justa y si la Dirección resolviera aceptar la obra, quedará el contratista obligado a conformarse con la rebaja acordada.

10. Cuando se juzgue necesario emplear materiales o ejecutar obras que no figuren en el presupuesto de la contrata, se evaluará su importe a los precios asignados a otras obras o materiales análogos si los hubiere y cuando no, se discutirán entre el Ingeniero Director y el contratista, sometiéndolos a la aprobación de la Dirección. Los nuevos precios convenidos por uno u otro procedimiento, se sujetarán siempre al establecido en el punto anterior.

11. Cuando el contratista, con autorización del Ingeniero Director de obras, emplee materiales de calidad más elevada o de mayores dimensiones de lo estipulado en el proyecto, o sustituya una clase de fabricación por otra que tenga asignado mayor precio o ejecute con mayores dimensiones cualquier otra parte de las obras, o en general, introduzca en ellas cualquier modificación que sea beneficiosa a juicio del Ingeniero Director de obras, no tendrá derecho sin embargo, sino a lo que le correspondería si hubiera realizado la obra con estricta sujeción a lo proyectado y contratado.

12. Las cantidades calculadas para obras accesorias, aunque figuren por partida alzada en el presupuesto final (general), no serán abonadas sino a los precios de la contrata, según las condiciones de la misma y los proyectos particulares que para ellas se formen, o en su defecto, por lo que resulte de su medición final.

13. El contratista queda obligado a abonar al Ingeniero autor del proyecto y director de obras así como a los Ingenieros Técnicos, el importe de sus respectivos honorarios facultativos por formación del proyecto, dirección técnica y administración en su caso, con arreglo a las tarifas y honorarios vigentes.

14. Concluida la ejecución de la obra, será reconocida por el Ingeniero Director que a tal efecto designe la empresa.

15. La garantía definitiva será del 4% del presupuesto y la provisional del 2%.

16. La forma de pago será por certificaciones mensuales de la obra ejecutada, de acuerdo con los precios del presupuesto, deducida la baja si la hubiera.

17. La fecha de comienzo de las obras será a partir de los 15 días naturales del replanteo oficial de las mismas y la definitiva, al año de haber ejecutado la provisional, procediéndose si no existe reclamación alguna, a la reclamación de la fianza.

18. Si el contratista al efectuar el replanteo, observase algún error en el proyecto, deberá comunicarlo en el plazo de quince días al Ingeniero Director de obras, pues transcurrido ese plazo será responsable de la exactitud del proyecto.

19. El contratista está obligado a designar una persona responsable que se entenderá con el Ingeniero Director de obras, o con el delegado que éste designe, para todo relacionado con ella. Al ser el Ingeniero Director de obras el que interpreta el proyecto, el contratista deberá consultarle cualquier duda que surja en su realización.

20. Durante la realización de la obra, se girarán visitas de inspección por personal facultativo de la empresa cliente, para hacer las comprobaciones que se crean oportunas. Es obligación del contratista, la conservación de la obra ya ejecutada hasta la recepción de la misma, por lo que el deterioro parcial o total de ella, aunque sea por agentes atmosféricos u otras causas, deberá ser reparado o reconstruido por su cuenta.

21. El contratista, deberá realizar la obra en el plazo mencionado a partir de la fecha del contrato, incurriendo en multa, por retraso de la ejecución siempre que éste no sea debido a causas de fuerza mayor. A la terminación de la obra, se hará una recepción provisional previo reconocimiento y examen por la dirección técnica, el depositario de

efectos, el interventor y el jefe de servicio o un representante, estampando su conformidad el contratista.

22. Hecha la recepción provisional, se certificará al contratista el resto de la obra, reservándose la administración el importe de los gastos de conservación de la misma hasta su recepción definitiva y la fianza durante el tiempo señalado como plazo de garantía. La recepción definitiva se hará en las mismas condiciones que la provisional, extendiéndose el acta correspondiente. El Director Técnico propondrá a la Junta Económica la devolución de la fianza al contratista de acuerdo con las condiciones económicas legales establecidas.

23. Las tarifas para la determinación de honorarios, reguladas por orden de la Presidencia del Gobierno el 19 de Octubre de 1961, se aplicarán sobre el denominado en la actualidad "Presupuesto de Ejecución de Contrata" y anteriormente llamado "Presupuesto de Ejecución Material" que hoy designa otro concepto.

Condiciones particulares

La empresa consultora, que ha desarrollado el presente proyecto, lo entregará a la empresa cliente bajo las condiciones generales ya formuladas, debiendo añadirse las siguientes condiciones particulares:

1. La propiedad intelectual de los procesos descritos y analizados en el presente trabajo, pertenece por entero a la empresa consultora representada por el Ingeniero Director del Proyecto.

2. La empresa consultora se reserva el derecho a la utilización total o parcial de los resultados de la investigación realizada para desarrollar el siguiente proyecto, bien para su publicación o bien para su uso en trabajos o proyectos posteriores, para la misma empresa cliente o para otra.

3. Cualquier tipo de reproducción aparte de las reseñadas en las condiciones generales, bien sea para uso particular de la empresa cliente, o para cualquier otra aplicación, contará con autorización expresa y por escrito del Ingeniero Director del Proyecto, que actuará en representación de la empresa consultora.

4. En la autorización se ha de hacer constar la aplicación a que se destinan sus reproducciones así como su cantidad.

5. En todas las reproducciones se indicará su procedencia, explicitando el nombre del proyecto, nombre del Ingeniero Director y de la empresa consultora.

6. Si el proyecto pasa la etapa de desarrollo, cualquier modificación que se realice sobre él, deberá ser notificada al Ingeniero Director del Proyecto y a criterio de éste, la empresa consultora decidirá aceptar o no la modificación propuesta.

7. Si la modificación se acepta, la empresa consultora se hará responsable al mismo nivel que el proyecto inicial del que resulta el añadirla.

8. Si la modificación no es aceptada, por el contrario, la empresa consultora declinará toda responsabilidad que se derive de la aplicación o influencia de la misma.

9. Si la empresa cliente decide desarrollar industrialmente uno o varios productos en los que resulte parcial o totalmente aplicable el estudio de este proyecto, deberá comunicarlo a la empresa consultora.

10. La empresa consultora no se responsabiliza de los efectos laterales que se puedan producir en el momento en que se utilice la herramienta objeto del presente proyecto para la realización de otras aplicaciones.

11. La empresa consultora tendrá prioridad respecto a otras en la elaboración de los proyectos auxiliares que fuese necesario desarrollar para dicha aplicación industrial, siempre que no haga explícita renuncia a este hecho. En este caso, deberá autorizar expresamente los proyectos presentados por otros.

12. El Ingeniero Director del presente proyecto, será el responsable de la dirección de la aplicación industrial siempre que la empresa consultora lo estime oportuno. En caso contrario, la persona designada deberá contar con la autorización del mismo, quien delegará en él las responsabilidades que ostente.