A speech production model including the nasal Cavity

A novel approach to articulatory analysis of speech signals.

Olesen, Morten

Publication date:
1995

Document Version
Tidlig version også kaldet pre-print

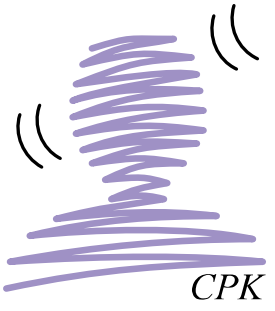
A Speech Production Model including the Nasal Cavity

A novel approach to articulatory analysis of speech signals

PhD Thesis

Morten Olesen

October 1995



Center for PersonKommunikation
Department of Communication Technology
Institute of Electronic Systems
Aalborg University
Denmark

ISBN 87-985750-0-7 ISSN 0908-1224 R 95-1007

En taleproduktionsmodel der inkluderer næsehulen

En ny indgangsvinkel til artikulatorisk analyse af talesignaler

Dansk resumé

For at opnå artikulatorisk analyse af talesignaler bliver den almindelige taleproduktionsmodel forbedret. Standard modellen, som den bruges i LPC analyse, modellerer i udstrakt grad kun akustiske egenskaber ved talen og er således ikke velegnet til artikulatorisk modellering og analyse. På trods af dette forhold er LPC modellen langt den mest brugte indenfor taleteknologi.

Ph.D. rapporten beskriver udvidelser af standard modellen på to punkter:

Næsehulen, der er en del af det menneskelige taleproduktionssystem, bliver inkluderet i modellen, og den modsvarende matematiske model, der indeholder både poler og nulpunkter, bliver etableret. Overføringsfunktionen bestemmes som et relativt kompliceret udtryk, men ved hjælp af et program til symbolsk matematik kan udtrykkets orden bestemmes. Når først ordenen er kendt, kan der benyttes systemidentifikationsmetoder til at finde overføringsfunktionen på normal form. Ud fra denne form kan polerne og nulpunkterne bestemmes.

Endvidere beskrives en algoritme til signal analyse, der modsvarer den udvidede taleproduktionsmodel. Algoritmen fjerner bidraget fra glottissignalet til talesignalet ved at estimere parametrene til en glottismodel. Når bidraget fra glottissignalet er fjernet, resterer bidraget fra talekanalen, som primært er bestemt at talekanalens form. Algoritmen er forbedret ved at tillade frekvensvægtning af fejlspektret uden samtidig at forringe analysens tidsmæssige opløsning.

De beskrevede delkomponenter bidrager til forskningen indenfor artikulatorisk taleanalyse, og er således ikke et forsøg på at finde den samlede endegyldige løsning på problemet.

A Speech Production Model including the Nasal Cavity

A novel approach to articulatory analysis of speech signals

Morten Olesen, PhD Thesis

October 1995

Abstract

In order to obtain articulatory analysis of speech signals the standard speech production model is improved. The standard model, as used in LPC analysis, to a large extent only models the acoustic properties of the speech signal as opposed to articulatory modelling of the speech production. In spite of this the LPC model is by far the most widely used model in speech technology.

The thesis presents research in which the standard model is enhanced in two respects:

Firstly the nasal cavity in the human speech production system is incorporated into the model and the corresponding mathematical model, which contains both poles and zeros, is established. The transfer function is determined as a fairly complex expression, but using a program for symbolic mathematics the order is determined. Once the order is known, system identification techniques are applied to determine the transfer function on normal form, from which the poles and zeros are obtainable.

Secondly a signal analysis algorithm corresponding to the extended speech production model is described. The algorithm extracts the glottal signal contribution from the speech signal by estimating the parameters of a glottal signal model, thereby obtaining the transfer function of the vocal tract. It is desired to determine the transfer function of the vocal tract isolated from excitation signal characteristics because the transfer function closely corresponds to the vocal tract shape. The algorithm is improved to allow frequency weighting of the error spectrum without sacrificing the time resolution of the analysis.

The described components contribute to the research in articulatory speech analysis rather than being an attempt to find a complete and final solution to the problem.

Olesen, M.: A speech production model including the nasal cavity - a novel approach to articulatory analysis of speech signals, 1995, Aalborg

ISBN 87-985750-0-7

ISSN 0908-1224

Institute internal registration number: R 95-1007

Center for PersonKommunikation

Department of Communication Technology

Institute of Electronic Systems

Aalborg University

Denmark

First printing May 16th 1995: 20 copies

Second revised printing November 16th 1995: 50 copies

Table of contents

List of figures

List of tables

Preface

Since the beginning of this study I have been fascinated by the lack of precise data describing the actual process of human speech production under natural circumstances. It never ceased to astonish me that after so many years of qualified research in phonetics and speech technology, no ideal method has been devised to obtain what to me seems to be one of the most fundamental descriptions of human speech production: the precise shape of the vocal tract as a function of time.

This thesis investigates possible improvements of previously used methods for articulatory analysis of speech signals in order to make the analysis applicable to nasal speech sounds and make the analysis more accurate in general.

Chapter 1 is a general introduction to the field of articulatory speech analysis, and at the end of the chapter the objective of the thesis is stated. In chapter 2 an enhanced speech production model is established which includes a model of the nasal cavity. In this chapter the mathematical equivalent of the physical model is found together with the transfer function and a method is developed to determine its poles and zeros. Chapter 3 discusses signal analysis algorithms suited for speech analysis based on the established speech production model.

The incorporation of the speech production model and the signal analysis algorithm into a complete articulatory speech analysis system is discussed in chapter 4. Among the appendices should be mentioned that a recorded speech corpus is documented in appendices C and D together with the associated signal processing to obtain speech data which are free of acoustic reflections and phase and amplitude distortions.

This Ph.D. study has been funded by a graduate scholarship (kandidatstipendiat) at the Faculty of Science and Technology at Aalborg University where I was employed for the first $2\frac{1}{2}$ years by the Communications Technology department within the Institute of Electronic Systems. The last part of the study was carried out concurrently to my employment as a research assistant in speech coding at Center for PersonKommunikation which is funded by the Danish Technical Research Council (STVF). During the whole study period I have fulfilled various forms of teaching obligations both internally and externally to the university.

I wish to thank my advisors Egon Hansen and Paul Dalsgaard for their support.

Morten Olesen Aalborg, May 1995

1 Articulatory speech analysis

This chapter introduces the research area of articulatory speech analysis in general, and in the last section the objective of this thesis is stated.

1.1 Introduction

Speech analysis is an integral part of all the fields of speech technology: recognition, coding, synthesis etc. and most speech analysis techniques have so far been based on Linear Predictive Coding (LPC), [Markel and Gray, 1976]. As will be discussed in section 1.3.3 this type of analysis results in very effective algorithms that only to a limited extent model the speech production process. In recent years however many of the subdisciplines in speech technology research have refined the algorithmic layers above the basic feature extraction to enable new methods of analysis by which at least some of the shortcomings in the model basis are revealed.

To augment the modelling capabilities in speech analysis one approach is to more precisely model the speech production process. An important objective of articulatory speech analysis is to extend the acoustic (spectral) modelling of the speech signal to incorporate a more detailed and complete description of the underlying speech production process - primarily the way in which the articulatory organs are moved during the production of speech.

1.2 Background and long term aim

The work described in this Ph.D. report is motivated by the fact that no databases exist for the shape of the human vocal tract during speech production [Schroeter and Sondhi, 1994]. This is an amazing fact since it is most likely that the vocal tract shape is one of the most fundamental sets of parameters in speech production modelling and therefore is believed to be very important in speech technology research in general. Given the importance of this data, it is evident that the absence of it is not caused by lack of interest or effort. Research groups in speech technology and phonetics have been dedicated to finding the exact vocal tract shape for several years, but only with partial success [Schroeder, 1967].

A database of vocal tract shapes is seen as an important long term aim of articulatory speech analysis. The database could consist of three time aligned sets of parameters as illustrated in figure 1-1:

- The vocal tract shape. This could be represented as the cross-sectional area as a function of distance from the glottis sampled spatially at sufficiently small intervals. Temporally the vocal tract shape should be represented so often that details of the movements of the articulatory organs would be revealed for all speech sounds.
- The speech signal resembling as close as possible the combined acoustic signal from the mouth and nostrils. This implies a high sample rate, linear phase and few or no acoustic reflections from the environment at recording time.
- Phonetic labels including suprasegmental markers (e.g. first and secondary stress) [Barry and Fourcin, 1992].

The database should cover all naturally occurring pronunciation variants statistically well. Once this has been achieved satisfactorily for a single speaker the methods can be applied to many speakers and languages. There are other important parameters to take into account in articulatory speech analysis but the most important aim is to obtain the vocal tract shape as described.

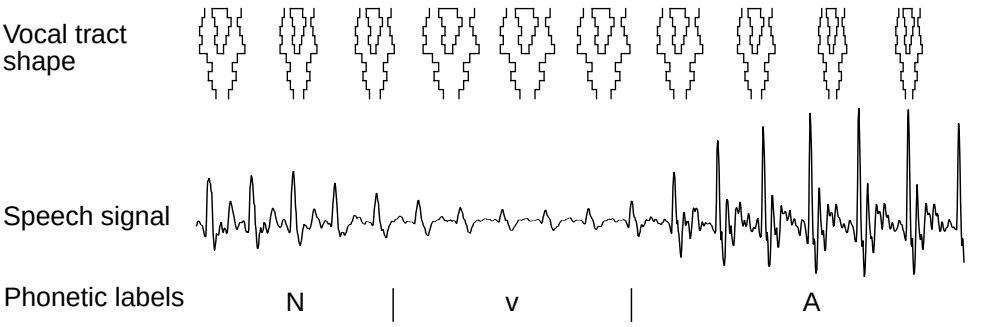


Figure 1-1: Illustration of the three time aligned sets of parameters which could constitute a vocal tract shape database (the vocal tract shapes are not actual).

A database of vocal tract shapes, their corresponding speech signals and annotation would as such be used in establishing a closer link between the two fields of 1) articulatory phonetics and 2) the knowledge of acoustic representations of speech from phonetics and speech technology in general. Each of the two fields have been studied extensively but with a contrast in emphasis: articulatory phonetics focuses on qualitative descriptions (e.g. place- and manner of articulation), general rules of articulation and especially coarticulation (e.g. [Kohler, 1990]). In contrast the field of acoustic descriptions is focused on quantitative measures like waveforms, spectra, durations, probability density functions and many other forms of statistics, but in general this field lacks rules of realisation of phonemes and coarticulation. Current practise shows that the articulatory domain knowledge is strong on rules and dependencies but weak on quantitative data and actual realizations, while the acoustic domain knowledge has the opposite strong and weak points. Improved models which are focused on the correlation between corresponding phenomena across the two domains would provide a way to utilize knowledge from one domain in the other.

1.3 Applications

In this section a number of possible applications are suggested for a vocal tract shape database as described in section 1.2. Since this evasive database does not yet exist, these possible applications should be read as part of the motivation for the work in establishing it.

1.3.1 Articulatory speech synthesis

One obvious area of application of a database of vocal tract shapes is within articulatory speech synthesis. Currently the work in this area focuses on turning the many qualitative rules of articulation and coarticulation into quantitative ones [Scully, 1987]. As shown in figure 1-1 these rules (1) transform a string of phonemes (a) into trajectories of articulatory parameters (b) from which the vocal tract shape (c) is determined using an articulatory model (2). The rules of

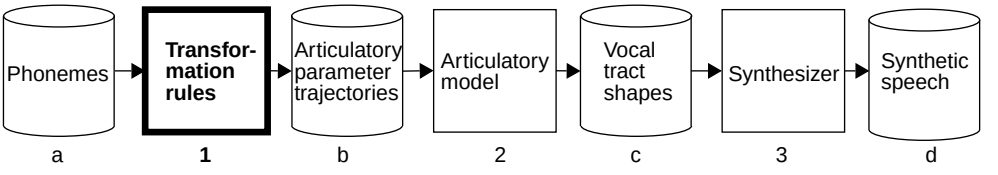


Figure 1-2: Parameters and transformations involved in articulatory speech synthesis. Time aligned data for modules (a) and (c) from a vocal tract shape database could assist in establishing the rules (1).

transformation are constructed using a great deal of phonetic knowledge in a trial-and-error process involving listening tests. A database of vocal tract shapes sequentially given as a function of time for given strings of phonemes would greatly facilitate and improve this process. Indeed it could be changed into a learning or statistical task where the rules were found semi-automatically based on a large database of *phonetic string – shape sequence* pairs.

In the database described on page 4 the articulatory parameters (b) are not included. In this case an invertible articulatory model could be used to invert the vocal tract shape to articulatory parameters in order to be able to construct the transformation rules (1). Alternatively the articulatory model could be included into the learned rules so that the rules would transform from phoneme sequences to vocal tract shapes directly (1+2).

The construction of articulatory models is currently most often based on a rudimentary knowledge of vocal tract shapes. From this a model is often constructed using simple geometric shapes (arcs, lines, splines etc.) in a mostly intuitive process [Coker, 1976]. However, it must be emphasised that what is modelled by this kind of articulatory model is the two–dimensional midsaggital cut of the vocal tract (a view of the vocal tract in a vertical plane as seen from the side and placed in the middle of the head). From this the cross-sectional areas are most often derived by multiplication by experimentally found coefficients. It is likely that the process of model construction using simple geometric shapes leads to an articulatory model which a) cannot model all vocal tract shapes as found in natural speech and b) is able to model shapes that do not occur in human speech production. Both of these points degrade the articula-

tory model which is intended to model the human speech production exactly. Given a nearly exhaustive database of vocal tract shapes, a basis would exist for the construction of articulatory models without the deficiencies just mentioned. Again, as with the transformation rules, a large number of quantitative data would allow statistical methods to be used in the construction of the models. One example of this method is the principle of constructing an articulatory model as a linear combination of principal components of vocal tract shapes. A method based on this has been applied successfully on (midsaggital) data extracted from X-ray film [Maëda, 1982] (see section 1.4.1).

1.3.2 Articulatory-phonetic feature estimation

Various approaches have been investigated on order to estimate parameters closely related to the use of the articulatory organs directly from a speech signal (e.g. acoustic-phonetic features [Dalsgaard, 1992]). Typically these systems must been trained without access to actual articulatory data. Although these approaches vary in the definition of the parameters to estimate, they would most likely benefit from a vocal tract shape database for training purposes, since target values are likely to be closely related to the vocal tract shape.

1.3.3 Speech production models

The increased insight into the physical process of speech production, such a database would give, could provide the information needed for the establishment of better speech production models. This is not only the case within articulatory speech synthesis as described in section 1.3.1. More generally speech production models serve as the core of most signal processing of speech signals and are thus fundamentally important to these.

Presently the almost exclusively used speech production model is the model used in linear predictive coding (LPC). In its simplicity this model has proven to be extremely well suited for modelling and parameterisation of certain classes of speech signals in the short term frequency domain. As described in section 1.4.3 the model also has an equivalency to a physical model of speech production. The LPC-model is well described in the literature [Markel and Gray, 1976], [Rabiner and Schafer, 1978]. However some of the known deficiencies of the LPC-model are:

- Lack of zeros in the transfer function
- No modelling of the excitation signal
- No physical modelling of the acoustic losses
- Short term time invariant modelling of a continuously time varying process

Some of these deficiencies could be corrected given a vocal tract shape database.

1.3.4 Articulatory phonetics

A vocal tract shape database could change articulatory phonetics substantially. So far only a few of the articulators have been accessible for direct measurement (e.g. lips and jaw), but the movements of most of the other articulators are not known in detail. As mentioned in the beginning of this chapter the phonetic research could be changed from being mostly qualitative in nature to being more quantitatively oriented. This would probably facilitate the integration of articulatory phonetics into speech technology (and vice versa).

1.4 Previous approaches

From the importance of the applications just described it is evident that a database of vocal tract shapes is and has long been a central desire amongst speech researchers. In this section the most important of the approaches taken so far for obtaining the data will be reviewed.

1.4.1 X-ray film

The profile of the articulators and their movements during speech production is recorded on cineradiographic (X-ray) film along with the recording of the audio signal. Typically the frame rate is 50 frames per second [Bothorel et al., 1986]. Each frame is analysed and at selected intervals (typically 5mm) the midsaggital distance (e.g. between the hard palate and the tongue) is measured. It should be emphasized that the relationship between the midsaggital distance and the cross-sectional area is not known very well [Perrier and Boë, 1989]. It has been empirically assumed that $A = \alpha d^{\beta}$ where A is the area and d is the distance. α and β are found ad hoc and vary along the vocal tract.

The amount of this kind of X-ray film data is limited and by no means exhaustive. This is related to some concern about the safety of the speakers involved. To obtain a sufficiently high temporal resolution (low exposure time) the X-ray radiation level has to be relatively high.

These disadvantages taken into account this method is nevertheless still one of the most important for the study of vocal tract shape related data.

1.4.2 Other measuring techniques

Pellet tracing Common for this group of methods is that a few small pellets are fixed to the articulators which are subjects to the study (typically the rear, mid and blade of the tongue). The positions of the pellets in the midsaggital plane are traced using various techniques: X-ray microbeam [Nadler et al., 1987], alternating magnetic field [Perkell and Cohen, 1986] etc. The analysis is relatively accurate and has a good time resolution, but the positions of the few pellets only give a coarse picture of the articulation, which can also be impeded to some degree by the pellets.

Electropalatography A grid of sensors (typically 8×8) is placed on the palate. Each of the sensors measure electrically whether the tongue makes contact with the palate at that specific point [Cohen and Perkell, 1986]. From this the location and shape of the constriction area can be deduced. This of course is only relevant for consonants since the tongue does not make contact with the palate during production of vowels. The method gives no information regarding the vocal tract shape for the unconstricted areas.

MRI or Magnetic Resonance Imaging exploits the unequal magnetic resonance characteristics of tissue and air to obtain a 3-dimensional image of the vocal tract shape [Foldvik et al., 1991], [Foldvik et al., 1993]. This is a very promising technique which is the first to provide true 3-dimensional time evolving measurements of the vocal tract. A crucial factor for this technique is the acquisition time versus accuracy. A short acquisition time is desired for sampling of the moving shape but also results in high noise levels and inaccuracy in the images. Presently many repetitions of the same short utterance must be filmed to give sufficient data for analysis. Variations in articulation between repetitions and lack of sharpness in the images due to short acquisition times result in inaccurate models.

1.4.3 Acoustic inversion

The work documented in this report is in the field of acoustic inversion. The principle of this group of methods is to analyse a speech signal and derive the vocal tract shape that was involved in the production of it [Schroeter and Sondhi, 1994]. A speech production model is inverted in the sense that a given acoustic signal is matched using the model and the physical equivalent part of the model is found, effectively inverting the process of speech production.

One way of obtaining the inversion is by the principle of analysis-by-synthesis [Parthasarathy and Coker, 1990]. The parameters of a speech production model (among which are the cross-sectional areas of the vocal tract) are varied according to a certain strategy. Then an error measure is calculated (normally defined in the frequency domain) between the synthetic speech from the model and the given real speech. The search strategy then attempts to minimize the error by iteratively updating the model parameters which ultimatively are taken as the result of the analysis-by-synthesis algorithm.

A classic example of the acoustic inversion method is the inversion of the LPC-model of speech production [Markel and Gray, 1976], [Rabiner and Schafer, 1978]. Under a few elementary assumptions (see section 2.2.1 on page 17) an LPC-model of order P has the physical equivalent of a chain of P cylindrical tube sections. The tube sections all have equal lengths and the individual cross-sectional areas can be derived from the filter coefficients. Although this simple procedure is based on a number of oversimplifications of the speech production process which degrades the results, it has been used for many years for articulatory speech analysis especially for vowels [Fant, 1960]. Two of the oversimplifications are the centre of focus of the work documented in this report:

- During speech production involving a lowered uvula (nasal– and nasalised sounds) the modelling of the vocal tract as a single chain of tube sections is fundamentally wrong. For these sounds an additional parallel chain of tubes modelling the nasal cavity should be included in the model.
- In the production of voiced speech the assumption in the LPC-model of spectrally white excitation is far from met. It has been shown by several that the acoustic wave above glottis has a spectrum which is not white [Fujisaki and Ljungqvist, 1986]. This has the undesired effect on the LPC analysis that it is not only the transmission part of speech production (corresponding to the vocal tract) that is modelled by the filter coefficients but also the spectrum of the excitation signal. If this effect is not eliminated (by an alternative modelling of the excitation signal) the cross-sectional areas found from the coefficients will be inaccurate.

An apparently positive side of using the standard LPC-model is the simplicity and the straightforward way of calculating the cross-sectional areas. In reality this is a deception since the problem of acoustic inversion is ill-posed because the articulatory→acoustic mapping is many-to-one (several articulatory configurations can result in virtually the same acoustic signal) and therefore has no unique inverse. The problem of selecting the correct solution out of many possible solutions is nontrivial but can be aided by continuity constraints in the articulatory domain on shape and rate of change of the area function [Schroeter and Sondhi, 1994].

1.5 Thesis objective

As should be evident from the earlier parts of this chapter, the problem of articulatory speech analysis is very complex. It has been indicated [Guérin, 1991] that possibly it may only be solved by applying a combination of techniques. This study is not an attempt to solve the problem as such. Rather some elements in an improved speech analysis method are proposed. As a starting point the work in this report has been limited to voiced speech.

The objective of this thesis is to investigate the possibility of overcoming two crucial shortcomings in the traditional method for articulatory analysis of speech signals.

The LPC based acoustic inversion technique is enhanced by the following elements:

- Modelling of nasal speech production by inclusion of a model of the nasal cavity into the speech production model.
- Modelling of the excitation signal for voiced speech production.

These elements have led to the two main research issues which are documented in this report: 1. establishment of an enhanced speech production model including a nasal cavity and 2. a signal analysis algorithm corresponding to this model. The outlines of these two research issues are given below.

1.5.1 Establishment of an enhanced speech production model

The establishment of an enhanced speech production model including the nasal cavity is based on a physical model consisting of tube sections with a Y-junction modelling the splitting at the uvula of the pharynx into the nasal cavity and the oral tract. This is treated in sections 2.1-2.3 where the corresponding time domain signal model is derived. This time domain system equivalency exploits exactly the same fundamental acoustic assumptions as in the equivalency between the mathematical LPC model and the corresponding physical model of tube sections which has been illustrated earlier (e.g. [Markel and Gray, 1976]).

The transfer function of the enhanced speech production model is derived in section 2.4. This derivation is non-trivial since all of the three branches of the physical model depend on each other. Consequently the transfer function is expressed as a number of subexpressions which in combination amount to a relatively complex expression.

In spite of the complexity it proves possible to determine the number of poles and zeros in the transfer function. This is shown in section 2.5 together with appendix A in which a program for symbolic mathematics is applied to the task.

Once the order of the transfer function has been determined, the transfer function can be determined on normal form for any given set of cross-sectional area values for the tube sections. This is accomplished using system identification techniques which is outlined in section 2.6. The poles and zeros of the transfer function can be determined from the normal form expression by numerical root solving techniques.

The overall result of establishing the enhanced speech production model in chapter 2 is that given a set of cross-sectional areas of the tube sections, the poles and zeros of the speech production model can be determined.

1.5.2 Signal analysis algorithms corresponding to the enhanced speech production model

In chapter 3 signal analysis algorithms corresponding to the enhanced speech production model are discussed.

As the signal analysis algorithm is considered to be sensitive to acoustic reflections from the recording environment as well as any phase or amplitude distortions, a speech corpus is recorded in an anechoic room. Furthermore the recordings are equalized with respect to phase and amplitude characteristics of the recording equipment, which are measured using MLSSA techniques. The recordings and equalization are documented in appendices C, D and E.

The signal analysis algorithm must analyse every speech signal segment for the number of poles and zeros corresponding to the order of the transfer function of the speech production model.

Furthermore the algorithm must incorporate a model of the glottal signal for voiced speech in order to remove the spectral contributions to the speech signal from the excitation. In this way the algorithm obtains an estimate of the contribution from the vocal tract shape to the speech signal. This estimate should be matched by the speech production model in order to find the vocal tract shape corresponding to the analysed speech signal segment. The type of algorithm chosen that incorporates an pole-zero analysis and an excitation model is the so-called GARMA[1] algorithm, which is described in section 3.1.

A modified GARMA algorithm, dubbed WGARMA, facilitates weighting of the error spectrum in the analysis thereby achieving better system identification

1. The inner loop of a GARMA analysis corresponds to an ARX analysis in system identification terminology [Ljung, 1987].

performance without sacrificing the algorithmic complexity significantly. This modified algorithm is derived and discussed in section 3.2.

The incorporation of the two elements proposed in chapters 2 and 3 into a complete articulatory speech analysis system is discussed in chapter 4.

2 Speech production model

In this chapter an enhanced speech production model including a nasal cavity will be established.

The principal elements in this chapter are as follows: under a few fundamental acoustic assumptions a physical model consisting of small tube sections has an equivalent discrete time system. The transfer function of this system is derived and the numbers of poles and zeros are determined. Finally the location of the singularities can be found by system identification of the time domain system. Some of the subjects covered in this chapter have been treated in a more compressed form in [Olesen, 1993].

2.1 Physical model

The transmission part (as opposed to the excitation part) of the human speech generation system is approximated by a physical model consisting of tube sections. This approximation is identical to the one involved in the equiva-

lency between an all-pole discrete time system and a single-tract model of speech production [Markel and Gray, 1976].

With the purpose of enhancing the speech production model to allow modelling of nasal and nasalized articulation, a chain of tube sections modelling the nasal tract is added to the single-chain model.

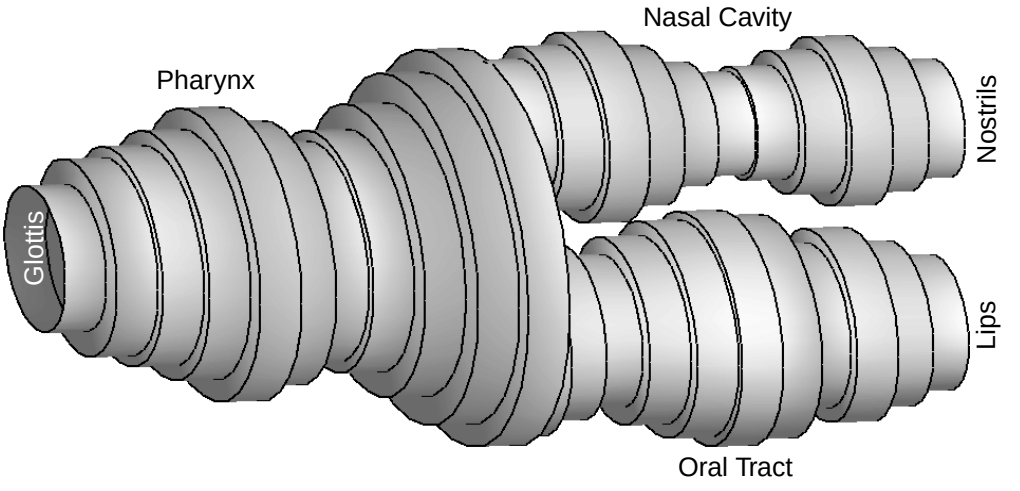


Figure 2-1: Physical speech production model enhanced by a nasal cavity. The specific diameters of the tube sections are more or less randomly chosen. The figure is meant to show the principle of connecting three chains of tube sections.

The resulting physical model, which is depicted in figure 2-1, then consists of three chains of tube sections which in turn model

- the pharynx (M_P sections)
- the oral tract above the uvula (M_O sections)
- the nasal cavity (M_N sections)

These three chains are connected at a **Y**-junction as shown in figure 2-2 where the cross-sectional area notation of the model is shown.

2.2 Mathematical model

This section outlines the derivation of the mathematical equivalent of the physical model enhanced by a nasal cavity. First the fundamental elements of the known equivalency between a single-chain tube model an all-pole mathematical model are briefly reviewed.

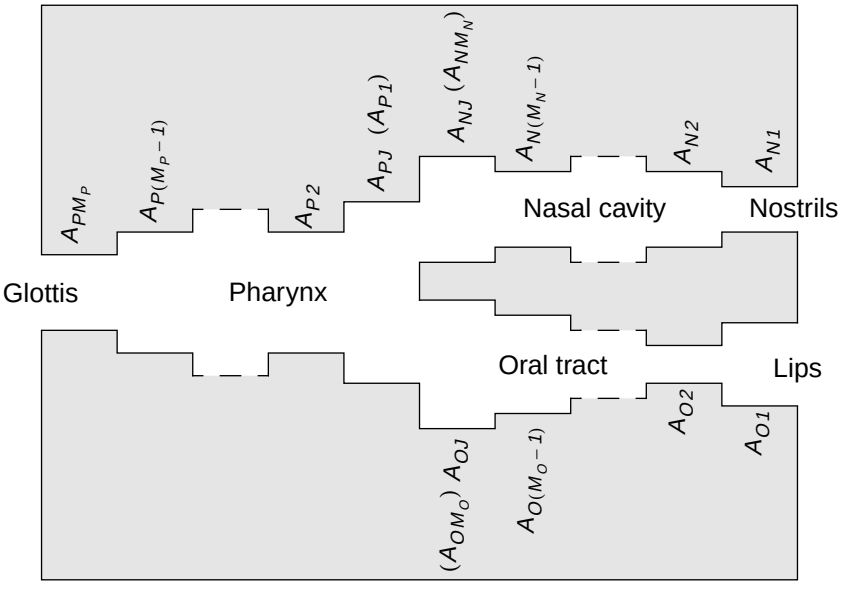


Figure 2-2: Cross-sectional view of the enhanced model showing section numbering. The sections at the **Y**-junction have two names to simplify some of the equations in this chapter.

2.2.1 A single tube section

Two elementary differential equations, known as the momentum equation (2.1) and the continuity of mass equation (2.2), describe the acoustic pressure and volume velocity in a single tube section [Markel and Gray, 1976], [Rabiner and Schafer, 1978].

$$\frac{\partial p_m(x,t)}{\partial x} = -\frac{\rho_0}{A_m}\frac{\partial u_m(x,t)}{\partial t} \tag{2.1}$$

$$\frac{\partial u_m(x,t)}{\partial x} = -\frac{A_m}{\rho_0 c^2}\frac{\partial p_m(x,t)}{\partial t} \tag{2.2}$$

where $u_m(x,t)$ and $p_m(x,t)$ are the acoustic volume velocity and pressure respectively at time t and distance x from the centre of the section (positive in the opposite direction of the glottis). c is the sound velocity and ρ_0 is the density of air.

For these equations to hold a few assumptions are made:

- the sound propagation can be viewed as plane wave, i.e. the wavelength is large compared to tube dimensions
- losses due to wall friction, vibration, viscosity, heat conduction etc. can be disregarded

Both of these assumptions are judged reasonable to make although not always completely fulfilled. In the nasal cavity the losses may be more important than elsewhere.

The solution to the equations (2.1)-(2.2) for the m'th tube section is [Markel and Gray, 1976]:

$$u_m(x, t) = u_m^+(t - x/c) - u_m^-(t + x/c) \tag{2.3}$$

$$p_m(x, t) = \frac{\rho_0 c}{A_m}(u_m^+(t - x/c) + u_m^-(t + x/c)) \tag{2.4}$$

The solution is interpreted as a linear combination of a forward travelling wave, u_m^+, and a reverse travelling wave, u_m^-.

2.2.2 Two adjacent tube sections

At the boundary between two sections there is continuity for both volume velocity and pressure, see figure 2-3 (λ is the section length):

$$u_m(\lambda/2, t) = u_{m-1}(-\lambda/2, t) \tag{2.5}$$

$$p_m(\lambda/2, t) = p_{m-1}(-\lambda/2, t) \tag{2.6}$$

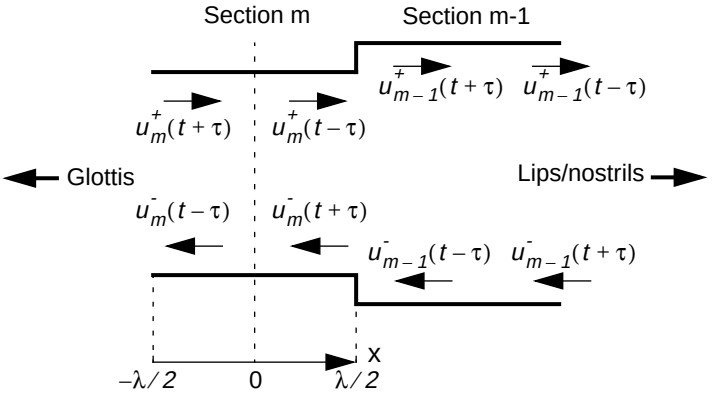


Figure 2-3: The volume velocity components in two adjacent tube sections.

Using equations (2.3)-(2.4) with τ defined as half the propagation time of a tube section ($\tau = \frac{\lambda}{2c}$):

$$u_m^+(t - \tau) - u_m^-(t + \tau) = u_{m-1}^+(t + \tau) - u_{m-1}^-(t - \tau) \tag{2.7}$$

$$u_m^+(t - \tau) + u_m^-(t + \tau) = \frac{A_m}{A_{m-1}}(u_{m-1}^+(t + \tau) + u_{m-1}^-(t - \tau)) \tag{2.8}$$

$u_m^-(t + \tau)$ is isolated from equation (2.8) and substituted into equation (2.7)

$$u_m^+(t-\tau) - \frac{A_m}{A_{m-1}}(u_{m-1}^+(t+\tau) + u_{m-1}^-(t-\tau)) + u_m^+(t-\tau) = u_{m-1}^+(t+\tau) - u_{m-1}^-(t-\tau) \Leftrightarrow \tag{2.9}$$

$$u_{m-1}^+(t+\tau) = \frac{2u_m^+(t-\tau) + \left(1 - \frac{A_m}{A_{m-1}}\right)u_{m-1}^-(t-\tau)}{1 + \frac{A_m}{A_{m-1}}} \Leftrightarrow \tag{2.10}$$

$$u_{m-1}^+(t+\tau) = \frac{2A_{m-1}}{A_{m-1}+A_m}u_m^+(t-\tau) + \frac{A_{m-1}-A_m}{A_{m-1}+A_m}u_{m-1}^-(t-\tau) \tag{2.11}$$

similarly $u_m^-(t+\tau)$ is found

$$u_m^-(t+\tau) = -\frac{A_{m-1}-A_m}{A_{m-1}+A_m}u_m^+(t-\tau) + \frac{2A_m}{A_{m-1}+A_m}u_{m-1}^-(t-\tau) \tag{2.12}$$

The reflection coefficients are defined as

$$\mu_m \doteq \frac{A_{m-1}-A_m}{A_{m-1}+A_m} \tag{2.13}$$

which from equations (2.11) and (2.12) result in the following system

$$u_{m-1}^+(t+\tau) = (1+\mu_m)u_m^+(t-\tau) + \mu_m u_{m-1}^-(t-\tau) \tag{2.14}$$

$$u_m^-(t+\tau) = -\mu_m u_m^+(t-\tau) + (1-\mu_m)u_{m-1}^-(t-\tau) \tag{2.15}$$

The junction between the two neighbouring sections described by equations (2.14)-(2.15) and the propagation delay of 2τ in section $m-1$ can be implemented as the system shown in figure 2-4.

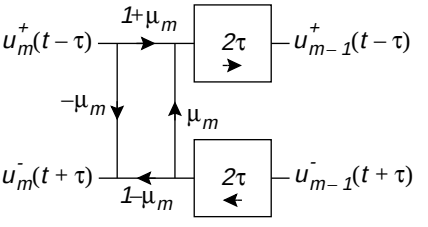


Figure 2-4: Discrete time system equivalent of a tube section and the transition to an adjacent section.

2.2.3 The Y-junction

The described methodology is applied at the **Y**-junction shown in figure 2-2.

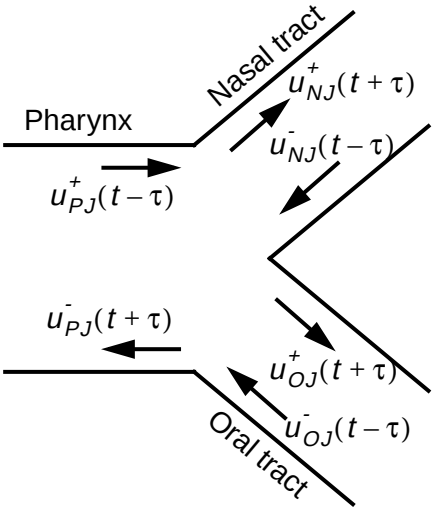


Figure 2-5: The volume velocity components near the **Y**-junction.

At the junction the continuity conditions are expressed as

$$u_{PJ}(\lambda/2, t) = u_{NJ}(-\lambda/2, t) + u_{OJ}(-\lambda/2, t) \quad (2.16)$$

$$p_{PJ}(\lambda/2, t) = p_{NJ}(-\lambda/2, t) = p_{OJ}(-\lambda/2, t) \quad (2.17)$$

This is analogous to equations (2.5) and (2.6).

Applying equation (2.3) to equation (2.16) gives

$$u^+_{PJ}(t-\tau) - u^-_{PJ}(t+\tau) = u^+_{NJ}(t+\tau) - u^-_{NJ}(t-\tau) + u^+_{OJ}(t+\tau) - u^-_{OJ}(t-\tau) \quad (2.18)$$

And using equation (2.4) equation (2.17) can be written as the two equations

$$u^+_{PJ}(t-\tau) + u^-_{PJ}(t+\tau) = \frac{A_{PJ}}{A_{NJ}}(u^+_{NJ}(t+\tau) + u^-_{NJ}(t-\tau)) \quad (2.19)$$

$$u^+_{OJ}(t+\tau) + u^-_{OJ}(t-\tau) = \frac{A_{OJ}}{A_{NJ}}(u^+_{NJ}(t+\tau) + u^-_{NJ}(t-\tau)) \quad (2.20)$$

$u^-_{PJ}(t+\tau)$ and $u^+_{OJ}(t+\tau)$ are isolated:

$$u^-_{PJ}(t+\tau) = \frac{A_{PJ}}{A_{NJ}}(u^+_{NJ}(t+\tau) + u^-_{NJ}(t-\tau)) - u^+_{PJ}(t-\tau) \quad (2.21)$$

$$u^+_{OJ}(t+\tau) = \frac{A_{OJ}}{A_{NJ}}(u^+_{NJ}(t+\tau) + u^-_{NJ}(t-\tau)) - u^-_{OJ}(t-\tau) \quad (2.22)$$

Equations (2.21) and (2.22) are substituted into equation (2.18):

$$u_{PJ}^{+}(t-\tau) - \frac{A_{PJ}}{A_{NJ}}(u_{NJ}^{+}(t+\tau) + u_{NJ}^{-}(t-\tau)) + u_{PJ}^{+}(t-\tau)$$
$$= u_{NJ}^{+}(t+\tau) - u_{NJ}^{-}(t-\tau) + \frac{A_{OJ}}{A_{NJ}}(u_{NJ}^{+}(t+\tau) + u_{NJ}^{-}(t-\tau) - 2u_{OJ}^{-}(t-\tau) \quad (2.23)$$

which by collecting terms with $u_{NJ}^{+}(t+\tau)$ on the left side yields:

$$u_{NJ}^{+}(t+\tau)\left(1 + \frac{A_{OJ}}{A_{NJ}} + \frac{A_{PJ}}{A_{NJ}}\right) \quad (2.24)$$
$$= u_{NJ}^{-}(t-\tau)\left(1 - \frac{A_{OJ}}{A_{NJ}} - \frac{A_{PJ}}{A_{NJ}}\right) + 2(u_{OJ}^{-}(t-\tau) + u_{PJ}^{+}(t-\tau))$$

$u_{NJ}^{+}(t+\tau)$ can be expressed using the definition (2.28):

$$u_{NJ}^{+}(t+\tau) \quad (2.25)$$
$$= \frac{2A_{NJ}}{A_{OJ}+A_{NJ}+A_{PJ}}(u_{PJ}^{+}(t-\tau) + u_{OJ}^{-}(t-\tau)) + \frac{A_{NJ}-A_{PJ}-A_{OJ}}{A_{OJ}+A_{NJ}+A_{PJ}}u_{NJ}^{-}(t-\tau)$$
$$= (1+\mu_{NJ})(u_{PJ}^{+}(t-\tau) + u_{OJ}^{-}(t-\tau)) + \mu_{NJ}u_{NJ}^{-}(t-\tau)$$

Similar expressions can be derived for $u_{OJ}^{+}(t+\tau)$ and $u_{PJ}^{-}(t+\tau)$:

$$u_{OJ}^{+}(t+\tau) = (1+\mu_{OJ})(u_{PJ}^{+}(t-\tau) + u_{NJ}^{-}(t-\tau)) + \mu_{OJ}u_{OJ}^{-}(t-\tau) \quad (2.26)$$

$$u_{PJ}^{-}(t+\tau) = (1+\mu_{PJ})(u_{NJ}^{-}(t-\tau) + u_{OJ}^{-}(t-\tau)) + \mu_{PJ}u_{PJ}^{+}(t-\tau) \quad (2.27)$$

where

$$\mu_{NJ} \doteq \frac{A_{NJ}-A_{PJ}-A_{OJ}}{A_{NJ}+A_{PJ}+A_{OJ}} \quad (2.28)$$

$$\mu_{OJ} \doteq \frac{A_{OJ}-A_{PJ}-A_{NJ}}{A_{NJ}+A_{PJ}+A_{OJ}} \quad (2.29)$$

$$\mu_{PJ} \doteq \frac{A_{PJ}-A_{NJ}-A_{OJ}}{A_{NJ}+A_{PJ}+A_{OJ}} \quad (2.30)$$

are called the reflection coefficients at the **Y**-junction.

The equations describing the junction, (2.25)-(2.30), can be implemented as the system in figure 2-6

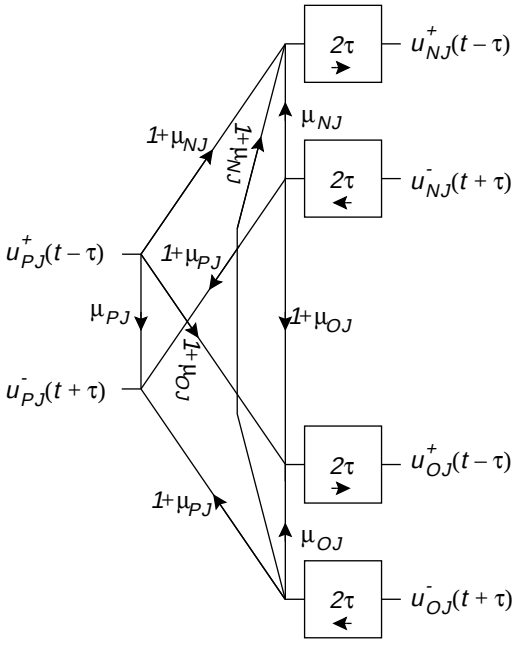


Figure 2-6: Mathematical equivalent of the **Y**-junction and the adjacent nasal and oral sections.

The reflection coefficients μ_{PG}, μ_{O1} and μ_{N1} which describe the acoustic coupling to the glottis and the radiation at the lips and nostrils are estimated in [Markel and Gray, 1976], [Rabiner and Schafer, 1978] as

$$\mu_{PG} = \frac{Z_G - \rho_0 c / A_{PM_P}}{Z_G + \rho_0 c / A_{PM_P}} \tag{2.31}$$

$$\mu_{O1} = \frac{Z_O - \rho_0 c / A_{O1}}{Z_O + \rho_0 c / A_{O1}} \tag{2.32}$$

$$\mu_{N1} = \frac{Z_N - \rho_0 c / A_{N1}}{Z_N + \rho_0 c / A_{N1}} \tag{2.33}$$

Where Z_G, Z_O and Z_N are the acoustic impedances at the glottis, lips and nostrils respectively. A detailed modelling would result in complex impedances, but here they are assumed real as is commonly seen.

The complete mathematical equivalent of the speech production model is shown in figure 2-7.

2.3 Discrete time implementation

If the number of sections in a chain of tubes is even then the impulse response from the system including any combination of reflections will be zero except at multipla of 4τ (twice the propagation time of a tube section). As a

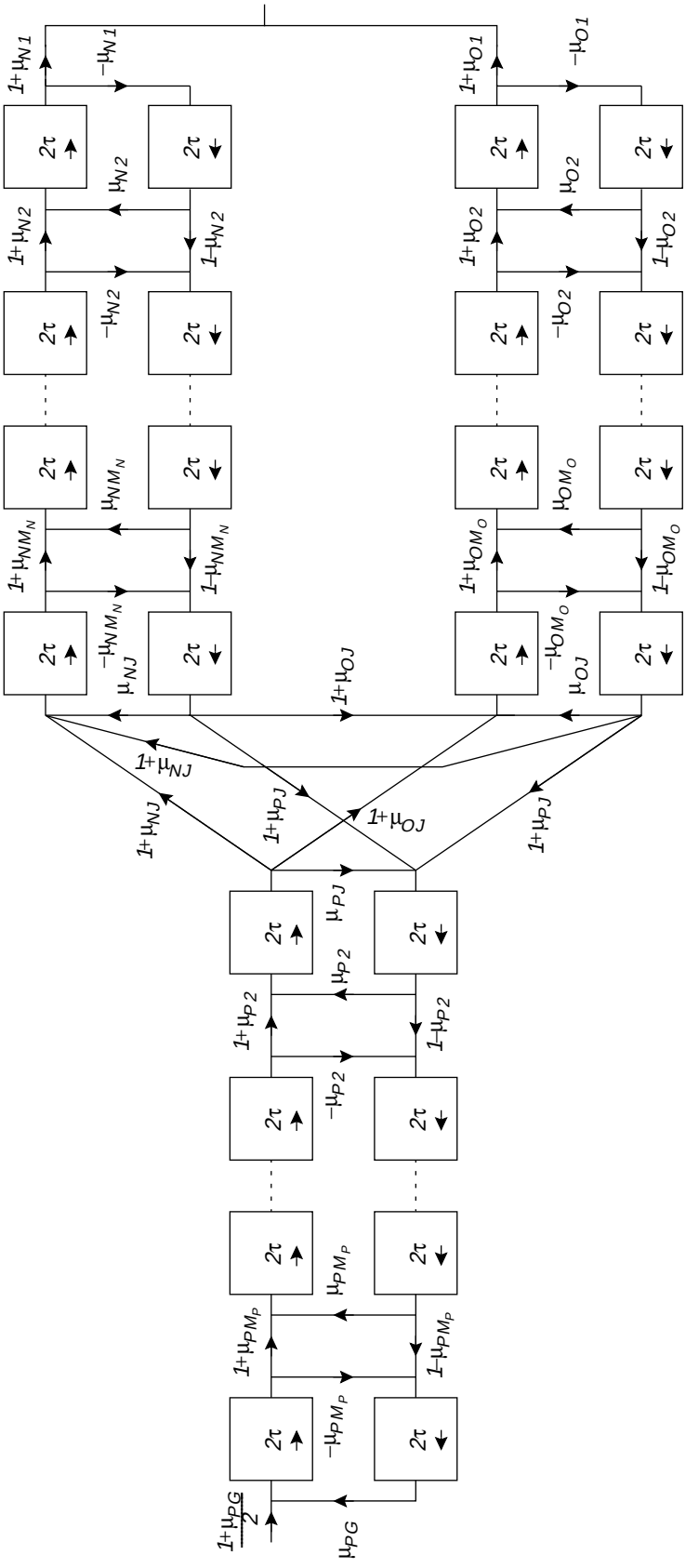


Figure 2-7: Mathematical equivalent of the enhanced speech production model.

consequence the sample time is chosen as 4τ along with the requirement that the number of tube sections in each tract must be even. In this case a sample frequency of 16 kHz is chosen which corresponds to a section length of 10.9 mm ($\tau = \frac{\lambda}{2c}$, $f_s = \frac{1}{4\tau}$, $c = 349\ m/s$).

Figure 2-8a shows the mathematical equivalent of to two tube sections and two transitions between sections. The variables $a(t), b(t), ..., f(t)$ denote the volume velocity at the indicated points.

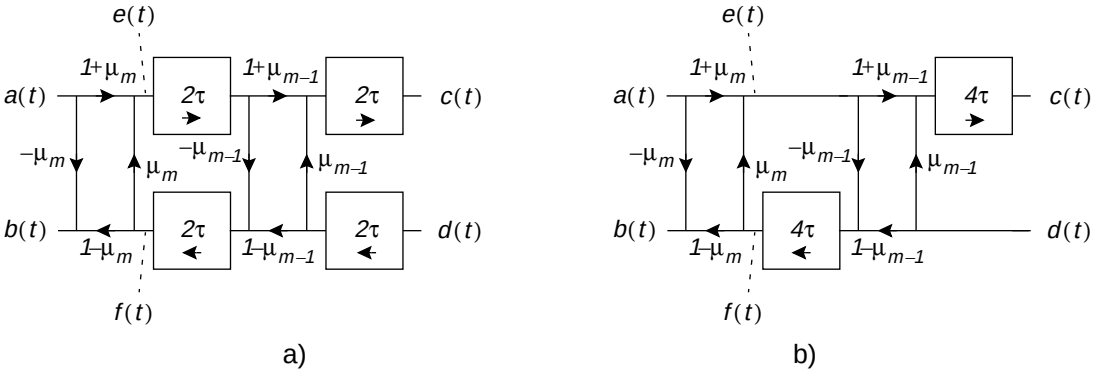


Figure 2-8: Discrete time system equivalent of two adjacent tube sections a) before moving the delays and b) after moving the delays according to equations (2.34)-(2.37).

At a sample time of 4τ a delay of 2τ corresponds to $z^{-1/2}$. To avoid this the signal graph in figure 2-8a is transformed according to the following equations:

$$e(t) = (1+\mu_m)a(t) + \mu_m(1-\mu_{m-1})d(t-4\tau) - \mu_m\mu_{m-1}e(t-4\tau) \quad (2.34)$$

$$c(t) = (1+\mu_{m-1})e(t-4\tau) + \mu_{m-1}d(t-4\tau) \quad (2.35)$$

$$f(t) = (1-\mu_{m-1})d(t-4\tau) - \mu_{m-1}(1+\mu_m)a(t-4\tau) - \mu_m\mu_{m-1}f(t-4\tau) \quad (2.36)$$

$$b(t) = (1-\mu_m)f(t) - \mu_m a(t) \quad (2.37)$$

From equations (2.34)-(2.37) it is apparent that the delays can be moved in such a way that only multiples of 4τ (corresponding to z^{-1}) appear which is shown in figure 2-8b.

If this principle of rearrangement is applied on the complete system (figure 2-7) a discrete time system suitable for z-transformation is obtained, see figure 2-9. After the rearrangement the internal values in the flow graph no longer represent the physical volume velocities as before, but rather the phase shifted values. Outside the transformation, however, the physical equivalency still holds (i.e. for U_G, U_N and U_O).

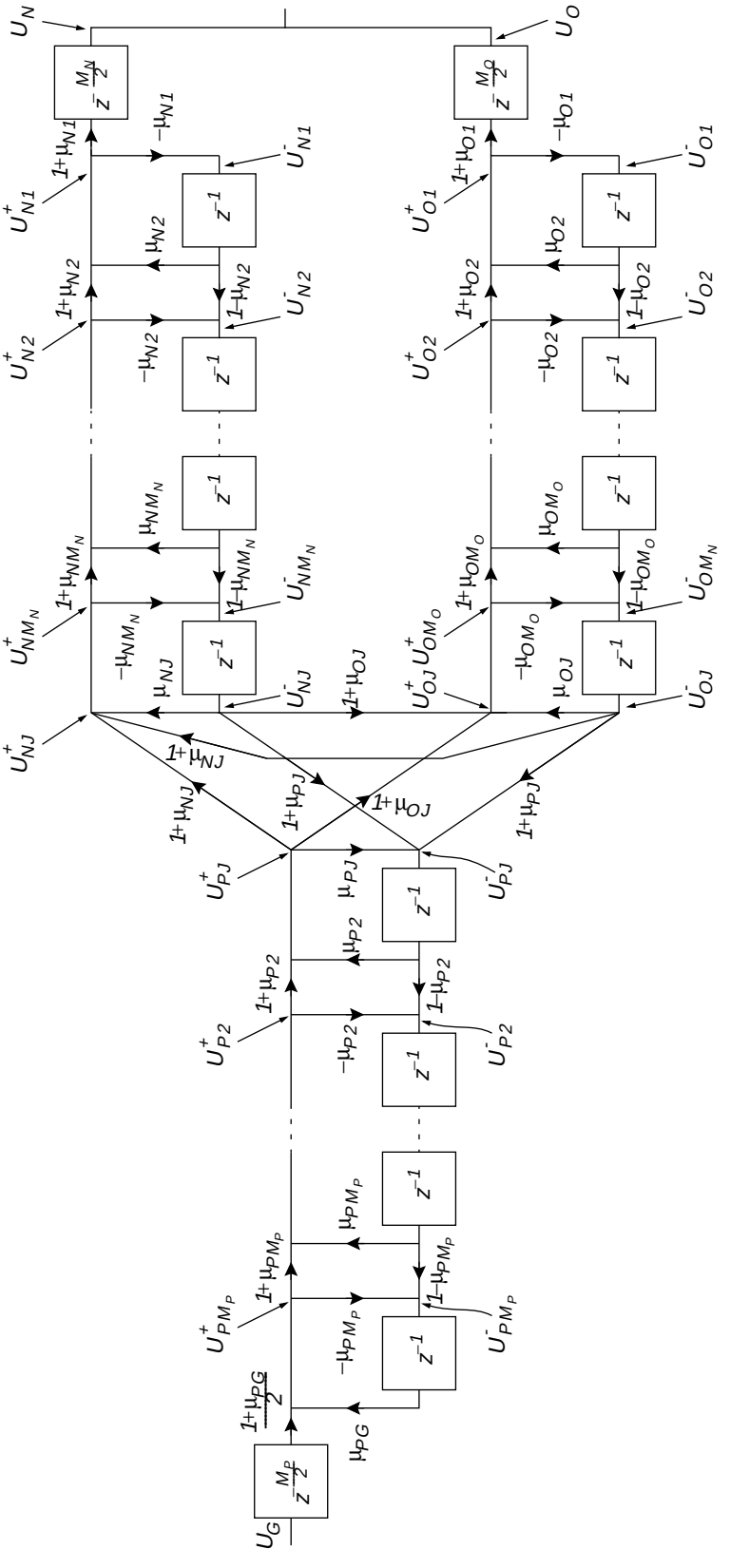


Figure 2-9: Discrete time system equivalent of the enhanced speech production model.

2.4 Transfer function by z-transform

In the calculation of the transfer function it is important to keep in mind that none of the three lattice sections in figure 2-9 can be regarded as an independent subsystem. Instead an expression is sought for the load that the nasal cavity and the oral tract exert as seen from the junction and the combined load as seen from the pharynx. These loads (subsequently called R_N, R_O and R_P respectively) are functions of z and can be interpreted as virtual reflection coefficients for each of the tracts.

The nasal tract is regarded first and from figure 2-9 it is seen that

$$U^{+}_{Nm-1} = (1+\mu_{Nm})U^{+}_{Nm} + \mu_{Nm}U^{-}_{Nm-1}z^{-1} \tag{2.38}$$

$$U^{-}_{Nm} = -\mu_{Nm}U^{+}_{Nm} + (1-\mu_{Nm})U^{-}_{Nm-1}z^{-1} \tag{2.39}$$

If equations (2.38) and (2.39) are rewritten as

$$U^{+}_{Nm} = \frac{1}{1+\mu_{Nm}}U^{+}_{Nm-1} - \frac{\mu_{Nm}}{1+\mu_{Nm}}z^{-1}U^{-}_{Nm-1} \tag{2.40}$$

$$U^{-}_{Nm} = \frac{-\mu_{Nm}}{1+\mu_{Nm}}U^{+}_{Nm-1} + \frac{1}{1+\mu_{Nm}}z^{-1}U^{-}_{Nm-1} \tag{2.41}$$

and by defining the 2 by 2 matrices

$$\boldsymbol{Q}_{Nm} \doteq \frac{1}{1+\mu_{Nm}}\begin{bmatrix} 1 & -\mu_{Nm}z^{-1} \\ -\mu_{Nm} & z^{-1} \end{bmatrix}, \quad 2 \le m \le M_N \tag{2.42}$$

and the column vectors

$$\boldsymbol{U}_{Nm} \doteq \begin{bmatrix} U^{+}_{Nm} \\ U^{-}_{Nm} \end{bmatrix} \tag{2.43}$$

then equations (2.40) and (2.41) can be written as

$$\boldsymbol{U}_{Nm} = \boldsymbol{Q}_{Nm}\boldsymbol{U}_{Nm-1} \tag{2.44}$$

Using the product of all the $\boldsymbol{Q}_N$'s

$$\boldsymbol{Q}_{NM_N}\cdots\boldsymbol{Q}_{N3}\boldsymbol{Q}_{N2} = \prod_{m=M_N}^{2}\boldsymbol{Q}_{Nm} \tag{2.45}$$

(which is also a 2 by 2 matrix) the nasal cavity can be described with

$$\boldsymbol{U}_{NM_N} = \prod_{m=M_N}^{2} \boldsymbol{Q}_{Nm} \; \boldsymbol{U}_{N1} \tag{2.46}$$

At the nostrils the volume velocity vector is found from figure 2-9

$$\boldsymbol{U}_{N1} = \begin{bmatrix} 1 \\ -\mu_{N1} \end{bmatrix} \frac{1}{1+\mu_{N1}} \; z^{M_N/2} U_N \tag{2.47}$$

And at the junction

$$\begin{bmatrix} U_{NJ}^{+} \\ zU_{NJ}^{-} \end{bmatrix} = \boldsymbol{U}_{NM_N} \tag{2.48}$$

The following expression is found for the nasal tract:

$$\begin{bmatrix} U_{NJ}^{+} \\ zU_{NJ}^{-} \end{bmatrix} = z^{M_N/2} \prod_{m=M_N}^{2} \boldsymbol{Q}_{Nm} \begin{bmatrix} 1 \\ -\mu_{N1} \end{bmatrix} \frac{U_N}{1+\mu_{N1}} \tag{2.49}$$

and the virtual reflection coefficient of the nasal tract is

$$R_N \doteq \frac{U_{NJ}^{+}}{U_{NJ}^{-}} = \frac{\begin{bmatrix} 1 & 0 \end{bmatrix} \prod_{m=M_N}^{2} \boldsymbol{Q}_{Nm} \begin{bmatrix} 1 \\ -\mu_{N1} \end{bmatrix}}{\begin{bmatrix} 0 & 1 \end{bmatrix} \prod_{m=M_N}^{2} \boldsymbol{Q}_{Nm} \begin{bmatrix} 1 \\ -\mu_{N1} \end{bmatrix}} z \tag{2.50}$$

where the row vector $\begin{bmatrix} 1 & 0 \end{bmatrix}$ simply picks the upper part of the following column vector. A similar expression exists for the virtual reflection coefficient for the oral tract R_O

$$R_O \doteq \frac{U_{OJ}^{+}}{U_{OJ}^{-}} = \frac{\begin{bmatrix} 1 & 0 \end{bmatrix} \prod_{m=M_O}^{2} \boldsymbol{Q}_{Om} \begin{bmatrix} 1 \\ -\mu_{O1} \end{bmatrix}}{\begin{bmatrix} 0 & 1 \end{bmatrix} \prod_{m=M_O}^{2} \boldsymbol{Q}_{Om} \begin{bmatrix} 1 \\ -\mu_{O1} \end{bmatrix}} z \tag{2.51}$$

From figure 2-9

$$U_{NJ}^{+} = (1+\mu_{NJ})(U_{PJ}^{+} + U_{OJ}^{-}) + \mu_{NJ} U_{NJ}^{-} \tag{2.52}$$

which is rearranged using equation (2.50)

$$(R_N - \mu_{NJ}) \frac{U_{NJ}^{-}}{U_{PJ}^{+}} = (1+\mu_{NJ}) \left(1 + \frac{U_{OJ}^{-}}{U_{PJ}^{+}} \right) \Leftrightarrow \tag{2.53}$$

$$H_D \doteq \frac{U_{NJ}^-}{U_{PJ}^+} = \frac{1+\mu_{NJ}}{R_N-\mu_{NJ}}\left(1+\frac{U_{OJ}^-}{U_{PJ}^+}\right) \tag{2.54}$$

From figure 2-9

$$U_{OJ}^+ = (1+\mu_{OJ})(U_{PJ}^+ + U_{NJ}^-) + \mu_{OJ}U_{OJ}^- \tag{2.55}$$

rearranging using R_O and H_D yields

$$(R_O-\mu_{OJ})\frac{U_{OJ}^-}{U_{PJ}^+} = (1+\mu_{OJ})\left(1+\frac{U_{NJ}^-}{U_{PJ}^+}\right) \Leftrightarrow \tag{2.56}$$

$$\begin{aligned} H_B \doteq \frac{U_{OJ}^-}{U_{PJ}^+} &= \frac{1+\mu_{OJ}}{R_O-\mu_{OJ}}\left(1+\frac{1+\mu_{NJ}}{R_N-\mu_{NJ}}\left(1+\frac{U_{OJ}^-}{U_{PJ}^+}\right)\right) \\ &= \frac{\frac{1+\mu_{OJ}}{R_O-\mu_{OJ}}\left(1+\frac{1+\mu_{NJ}}{R_N-\mu_{NJ}}\right)}{1-\frac{1+\mu_{OJ}}{R_O-\mu_{OJ}}\frac{1+\mu_{NJ}}{R_N-\mu_{NJ}}} \\ &= \frac{(1+\mu_{OJ})(R_N+1)}{(R_O-\mu_{OJ})(R_N-\mu_{NJ})-(1+\mu_{OJ})(1+\mu_{NJ})} \end{aligned} \tag{2.57}$$

Doing the same operation on equation (2.27) using H_D and H_B gives R_P

$$R_P \doteq \frac{U_{PJ}^-}{U_{PJ}^+} = \mu_{PJ} + (1+\mu_{PJ})(H_D+H_B) \tag{2.58}$$

The expression for the pharynx analogous to equation (2.46) is

$$\boldsymbol{U}_{PM_P} = \prod_{m=M_P}^{2} \boldsymbol{Q}_{Pm}\ \boldsymbol{U}_{PJ} \tag{2.59}$$

At the glottal termination:

$$U_G = \frac{2}{1+\mu_{PG}}\, z^{M_P/2}\begin{bmatrix}1 & -\mu_{PG}z^{-1}\end{bmatrix}\ \boldsymbol{U}_{PM_P} \tag{2.60}$$

Combining equations (2.58)-(2.60) gives

$$H_A \doteq \frac{U_G}{U_{PJ}^+} = \frac{2}{1+\mu_{PG}}\begin{bmatrix}1 & -\mu_{PG}z^{-1}\end{bmatrix} z^{M_P/2}\prod_{m=M_P}^{2}\boldsymbol{Q}_{Pm}\begin{bmatrix}1\\ R_P\end{bmatrix} \tag{2.61}$$

From equation (2.49):

$$H_E \doteq \frac{\bar{U}_{NJ}}{U_N} = \begin{bmatrix} 0 & 1 \end{bmatrix} z^{M_N/2-1} \prod_{m=M_N}^{2} \boldsymbol{Q}_{Nm} \begin{bmatrix} 1 \\ -\mu_{N1} \end{bmatrix} \frac{1}{1+\mu_{N1}} \tag{2.62}$$

For the oral tract a similar expression exists:

$$H_C \doteq \frac{\bar{U}_{OJ}}{U_O} = \begin{bmatrix} 0 & 1 \end{bmatrix} z^{M_O/2-1} \prod_{m=M_O}^{2} \boldsymbol{Q}_{Om} \begin{bmatrix} 1 \\ -\mu_{O1} \end{bmatrix} \frac{1}{1+\mu_{O1}} \tag{2.63}$$

Now all elements in the transfer function have been found

$$H = \frac{U_N + U_O}{U_G} = \frac{H_B}{H_A H_C} + \frac{H_D}{H_A H_E} \tag{2.64}$$

2.4.1 Evaluation of the transfer function

In this section the transfer function is evaluated for the arbitrarily chosen areas in table 2-1 and $\mu_{PG} = 0.7383$, $\mu_{O1} = 0.9810$ and $\mu_{N1} = 0.9873$.

Section number	1	2	3	4	5	6	7	8
Pharynx	2	4	8	13	14	9	5	3
Oral tract	3	5	9	13	13	10	6	4
Nasal cavity	2	5	8	14	12	9	4	2

Table 2-1: Areas used in the evaluation of the transfer function.

Figure 2-10 illustrates the dimensions corresponding to table 2-1.

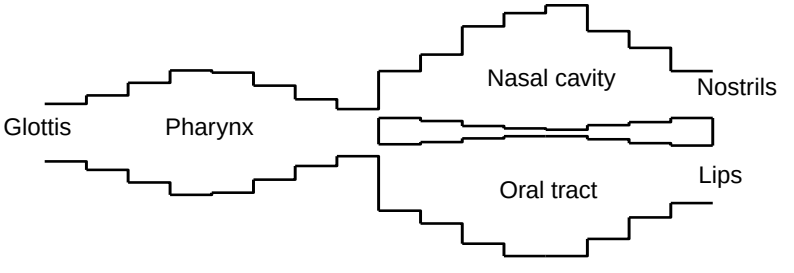


Figure 2-10: Shape of vocal tract corresponding to the areas in table 2-1.

Figure 2-11 shows the magnitude of the transfer function given these areas. An FFT of the impulse response of the time domain implementation of the system (figure 2-9) gives exactly the same curve, which is also the case for the phase. This indicates that the expression for the transfer function is correct.

The equations in this section express the transfer function in the z-domain which can be evaluated for given cross-sectional areas of the tube sections in

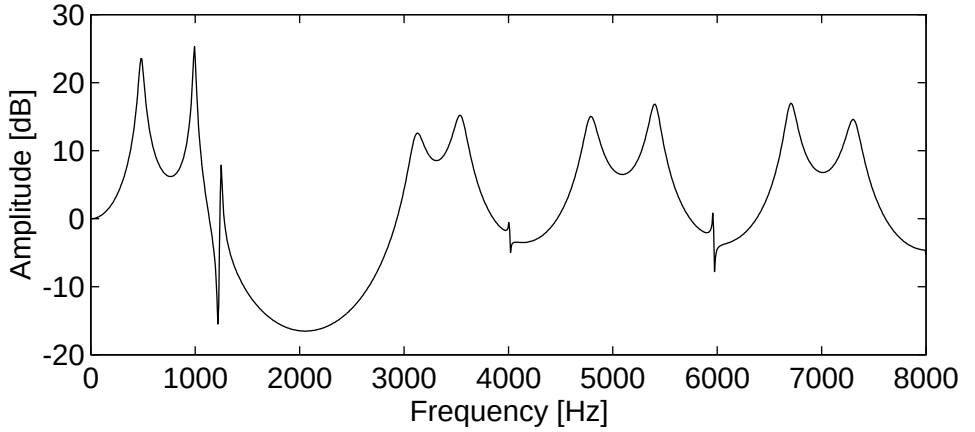


Figure 2-11: Magnitude transfer function as calculated by equation (2.64) at a sample frequency of 16 kHz. The presence of both poles and zeros is evident.

the pharynx, nasal cavity and oral tract together with termination impedances (μ_G, μ_{N1}, μ_{O1}), see figure 2-12. The surface plot illustrates that the transfer function can be evaluated for any given z.

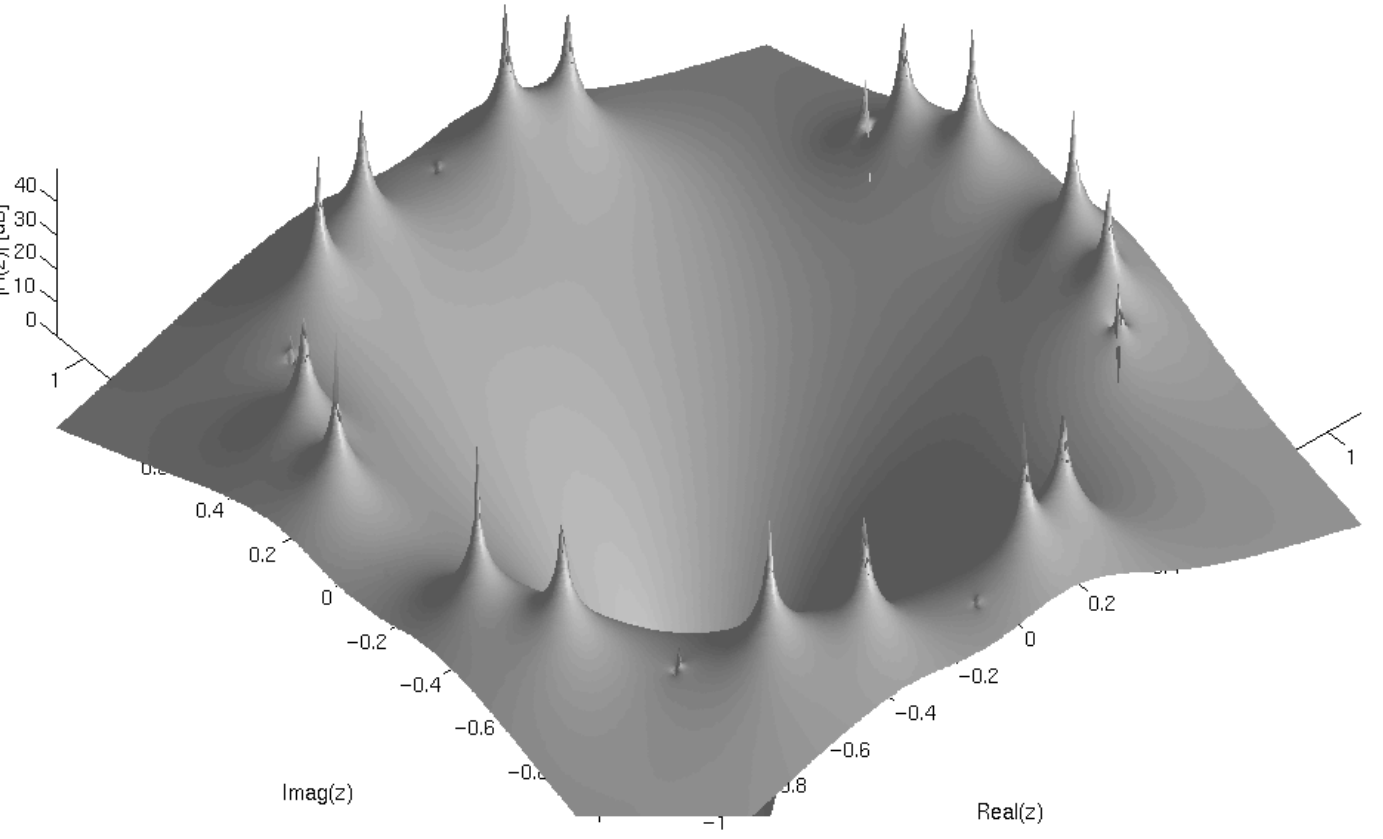


Figure 2-12: Surface plot of the magnitude of the transfer function as calculated by equation (2.64).

Although this has many applications and represents an improvement in modelling accuracy over the all-pole model, it is true that the expression is rather complex and in some respects not straightforward to apply. Most prominently the singularities, which are evident in figure 2-12, cannot be determined directly, since the transfer function is not expressed as a rational function, i.e. a ratio of polynomials in z [Oppenheim and Schafer, 1975]. In fact two sets of poles and zeros are so close (but not exactly equal) that they do not show up in the figure.

The following sections describe a method to determine the transfer function on the rational function form more directly from the cross-sectional areas.

2.5 Order of the transfer function

In most applications it is desirable to know the singularities of the transfer function. It has, however, due to the joint complexity of equations (2.50)-(2.64) not been possible to reduce the transfer function to its rational function form (equation (2.65)) by manipulation of the equations.

$$H(z) = \frac{\sum_{j=0}^{B} b_j z^{-j}}{1 + \sum_{i=1}^{A} a_i z^{-i}} \tag{2.65}$$

As will be shown in this section, however, it is possible to determine the order of the transfer function i.e. the values of A and B given the number of tube sections in each tract (M_P, M_N and M_O).

A common subexpression in equations (2.50), (2.51), (2.61)-(2.63) is a product of $M-1$ $\boldsymbol{Q}_m$-matrices. The elements of the matrix product are polynomials in z^{-1} of an order which shall be determined next.

If $X_i(n)$ is a polynomial in z^{-1} of order n

$$X_i(n) = \sum_{j=0}^{n} x_{ijn} z^{-j} \tag{2.66}$$

with x_{ijn} independent of z then a $\boldsymbol{Q}_m$-matrix (see equation (2.42)) can be written as

$$\boldsymbol{Q}_m = \frac{1}{1+\mu_m} \begin{bmatrix} 1 & -\mu_m z^{-1} \\ -\mu_m & z^{-1} \end{bmatrix} = \begin{bmatrix} X_1(0) & X_2(0) z^{-1} \\ X_3(0) & X_4(0) z^{-1} \end{bmatrix} \tag{2.67}$$

and the product of two $\boldsymbol{Q}$-matrices will be on the form

$$\begin{aligned}\boldsymbol{Q}_{m+1}\boldsymbol{Q}_m &= \frac{1}{1+\mu_{m+1}}\begin{bmatrix}1 & -\mu_{m+1}z^{-1}\\ -\mu_{m+1} & z^{-1}\end{bmatrix}\frac{1}{1+\mu_m}\begin{bmatrix}1 & -\mu_m z^{-1}\\ -\mu_m & z^{-1}\end{bmatrix}\\ &= \frac{1}{1+\mu_{m+1}}\frac{1}{1+\mu_m}\begin{bmatrix}1+\mu_{m+1}\mu_m z^{-1} & -\mu_m z^{-1}-\mu_{m+1}z^{-2}\\ -\mu_{m+1}-\mu_m z^{-1} & \mu_{m+1}\mu_m z^{-1}+z^{-2}\end{bmatrix}\\ &= \begin{bmatrix}X_1(1) & X_2(1)z^{-1}\\ X_3(1) & X_4(1)z^{-1}\end{bmatrix}\end{aligned} \tag{2.68}$$

The coefficients x_{ij1} can be determined, but in this respect the values are irrelevant. The relevant information from equation (2.68) is that the elements in the resulting product of two $\boldsymbol{Q}_m$-matrices are polynomials in z^{-1} of the order 1.

Generally multiplying a $\boldsymbol{Q}_m$-matrix with a product matrix with elements of the order n yields a new product matrix with elements of the order $n+1$. This is seen from equation (2.69):

$$\begin{aligned}\boldsymbol{Q}_m\begin{bmatrix}X_1(n) & X_2(n)z^{-1}\\ X_3(n) & X_4(n)z^{-1}\end{bmatrix} &= \frac{1}{1+\mu_m}\begin{bmatrix}X_1(n)-\mu_m z^{-1}X_3(n) & z^{-1}X_2(n)-\mu_m X_4(n)z^{-2}\\ -\mu_m X_1(n)+z^{-1}X_3(n) & -\mu_m z^{-1}X_2(n)+X_4(n)z^{-2}\end{bmatrix}\\ &= \begin{bmatrix}X_1(n+1) & X_2(n+1)z^{-1}\\ X_3(n+1) & X_4(n+1)z^{-1}\end{bmatrix}\end{aligned} \tag{2.69}$$

Again the values of the coefficients x_{ijn+1} are irrelevant

Combining these two results gives the rule that the product of N $\boldsymbol{Q}_m$-matrices is a product matrix with elements of the order $N-1$.

So products of $M-1$ $\boldsymbol{Q}_m$-matrices (which are common in equations (2.50), (2.51) and (2.61)-(2.63)) have elements which are polynomials in z^{-1} of order $M-2$.

For the nasal cavity equation (2.45) can be expressed as

$$\prod_{m=M_N}^{2}\boldsymbol{Q}_{Nm} \doteq \begin{bmatrix}N_1 & N_2 z^{-1}\\ N_3 & N_4 z^{-1}\end{bmatrix} \tag{2.70}$$

where the N's are polynomials in z^{-1} of order M_N-2 that in general have unequal coefficients. Similar equations exist for the pharynx and upper oral tract in which N/N is replaced by P/P and O/O respectively.

Programs for symbolic mathematics (primarily MapleV) were given the equations (2.50)-(2.64) and (2.70) and set up to simplify the transfer function. In appendix A on page 65 the result is shown and reduced to the form

$$H(z) = -\frac{1}{2}z^{4-M_P/2}\frac{\left(\begin{array}{l}(z^2O_1 - z(\mu_{O1}O_2 - O_3) - \mu_{O1}O_4)k_N z^{-M_N/2} \\ +(z^2N_1 - z(\mu_{N1}N_2 - N_3) - \mu_{N1}N_4)k_O z^{-M_O/2}\end{array}\right)}{f_6z^6 + f_5z^5 + f_4z^4 + f_3z^3 + f_2z^2 + f_1z + f_0} \quad (2.71)$$

where

$$k_N = (1+\mu_{PG})(1+\mu_{N1})(1+\mu_{NJ}) \quad (2.72)$$

$$k_O = (1+\mu_{PG})(1+\mu_{O1})(1+\mu_{OJ}) \quad (2.73)$$

and the f_i's are linear combinations of products of three polynomials, one for each tract. As an example are given the relatively few terms of f_5

$$\begin{aligned} f_5 = {} & P_1N_1O_2\mu_{O1} + P_1N_3O_1\mu_{NJ} - P_2N_1O_1\mu_{PJ} \\ & + P_3N_1O_1\mu_{PG} + P_1N_2O_1\mu_{N1} + P_1N_1O_3\mu_{OJ} \end{aligned} \quad (2.74)$$

As shown in the appendix the f_i's are comprehensive but they all share the property of being polynomials in z^{-1} of the order $M_P + M_N + M_O - 6$ (henceforth written as $poly(-(M_P + M_N + M_O - 6))$). Thus if the denominator of equation (2.71) is multiplied by z to the power of $M_P + M_N + M_O - 6$, it becomes a polynomial in z of order $M_P + M_N + M_O$ i.e. $poly(M_P + M_N + M_O)$.

Considering only the order of the numerator of equation (2.71) (likewise after multiplication by z to the power of $M_P + M_N + M_O - 6$) it can be written as

$$\begin{aligned} & z^{4-\frac{M_P}{2}+M_P+M_O+M_N-6}\left(\begin{array}{l}(z^2+z+1)poly(-(M_O-2))z^{-M_N/2} \\ +(z^2+z+1)poly(-(M_N-2))z^{-M_O/2}\end{array}\right) \\ & = z^{M_P/2}(poly(M_O)z^{M_N/2} + poly(M_N)z^{M_O/2}) \end{aligned} \quad (2.75)$$

So considering the order of $H(z)$, it can written as

$$H(z) = z^{M_P/2}\frac{poly(M_O)z^{M_N/2} + poly(M_N)z^{M_O/2}}{poly(M_P + M_N + M_O)} \quad (2.76)$$

From equation (2.76) the order of the transfer function can be determined:

As an example a model with 8 sections in each of the three tracts will have 24 poles, 8 zeros at origo and 8 other zeros.

As mentioned earlier simplification of the transfer function is not practicable in general. However, if the number of sections in each tract are chosen and specific values are given to the areas, the highly complex symbolic coefficients of

Singularities	Number
Poles	$M_P + M_N + M_O$
Zeros at 0	$\frac{1}{2}(M_P + min(M_O, M_N))$
Other zeros	$max(M_O, M_N)$

Table 2-2: Order of the transfer function.

z reduce to numbers. In this case it is indeed practicable with a symbolic mathematics program to determine the rational function form of the transfer function and thereby the order. This approach confirms the general rules given in table 2-2.

2.6 Transfer function by LS-analysis

This section outlines a method for determination of the transfer function at rational function form given specific areas for the tube sections in the three tracts.

The procedure is to generate an excitation signal, feed it through the time domain implementation of the discrete time system (figure 2-9) and perform a least-squares system identification on the response to obtain the a_i's and b_j's of equation (2.65) [Fujisaki and Ljungqvist, 1987], [Wang et al., 1990].

Since the system order is known from table 2-2 and no noise except round off error is introduced in the calculations, the least squares prediction error is virtually zero and the accuracy of the coefficients found is limited only by the precision of the representation of numbers in the specific computer and program. Gaussian noise was used as excitation signal. Figure 2-13 illustrates the poles and zeros of the exemplary system as calculated by this method. The plot is directly comparable with figure 2-12.

The system identification method involves calculation of autocorrelation and cross-correlation coefficients for the excitation and response signals and inversion of a symmetric matrix which can be done effectively with the Cholesky decomposition method.

The algorithm used for the system identification is also used for speech signal analysis and is described in chapter 3.

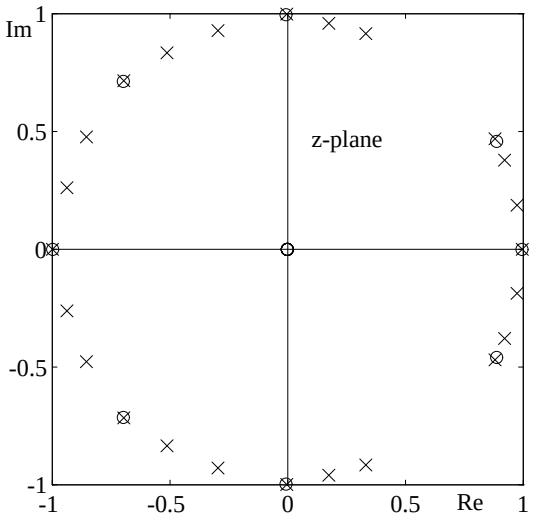


Figure 2-13: Poles and zeros of the transfer function calculated by system identification of the time domain system. This figure may be compared to figure 2-12.

2.7 Summary

In this chapter a speech production model has been established which is enhanced by a model of the nasal cavity compared the standard model used in LPC analysis.

It is not claimed that the speech production model is a perfect model of the human speech production. On the contrary some known factors in speech production, such as acoustic losses, are not modelled. But with respect to nasalized speech production, this new speech production model is superior.

Seen on a system level the function of the speech production model is to determine the poles and zeros in the transfer function given the cross-sectional areas of the tube sections in the model. The model is not directly invertible and indeed multiple vocal tract shapes may result in approximately the same poles and zeros.

3 Signal analysis algorithm

In this chapter a signal analysis algorithm is considered which corresponds to the speech production model described in the previous chapter. As a first step the algorithm is limited to voiced speech.

In speech production theory the short-time spectrum of a voiced speech signal is commonly considered as composed of three components [Fant and Lin, 1988]:

a) The spectrum of the excitation signal at the glottis

b) The transfer function of the physical system consisting of the pharynx and oral- and nasal cavities

c) The spectral characteristics of the radiation from the lips and nostrils

The assumption is made, that b) is determined predominantly by the vocal tract shape. For that reason the contributions from a) and c) are required to be removed from the speech signal in order to achieve articulatory analysis. This is accomplished for each speech frame by simultaneously estimating the

parameters of 1) a glottal signal/radiation model and 2) the transfer function of the speech production model.

In this case the purpose of modelling the glottal signal and radiation characteristics is to remove these components from the short time spectrum of the speech signal in order to obtain the transfer function component. Since the glottal signal is unknown and constantly changing during speech production, the method adopted is to estimate the glottal signal for every speech frame together with the transfer function of the speech production model.

Since the speech production model described in chapter 2 has poles and zeros in the transfer function, the corresponding signal analysis algorithm must be of the ARMA[1] type. For given numbers of tube sections in the pharynx and oral- and nasal cavities the number of poles and zeros in the transfer function are known from table 2-2 on page 34. This is also the order of the ARMA-analysis.

3.1 GARMA analysis

An ARMA-analysis algorithm incorporating a model of the glottal signal has been proposed earlier [Fujisaki and Ljungqvist, 1987]. The term *GARMA* is assigned to this kind of analysis as a mnemonic for Glottal ARMA. In this section the mathematics behind the analysis is described. In section 3.2 modified versions of the algorithm are discussed.

Initially the analysis carried out to obtain the filter coefficients in the transfer function for a known excitation signal is described. After that the iterative procedure for estimation of both the excitation signal and the transfer function is introduced.

3.1.1 Estimation of the transfer function for a given excitation signal

Assuming two known signals $s(n)$ and $g(n)$ representing the speech signal and the excitation signal respectively. A prediction model serves as basis for the algorithm

$$\hat{s}(n) = \sum_{j=0}^{q} b_j g(n-j) - \sum_{i=1}^{p} a_i s(n-i) \qquad (3.1)$$

where $\hat{s}(n)$ is the predicted speech signal. In order to clarify the nature of the prediction model, it is z-transformed

1. See footnote on page 12.

$$\hat{S}(z) = B(z)G(z) - S(z)(A(z) - 1) \tag{3.2}$$

where

$$A(z) = a_0 + a_1 z^{-1} + \ldots + a_p z^{-p} \quad , \quad a_0 = 1 \tag{3.3}$$

$$B(z) = b_0 + b_1 z^{-1} + \ldots + b_q z^{-q} \tag{3.4}$$

If the model accurately predicts the speech i.e. $\hat{s}(n) = s(n)$ then

$$\hat{S}(z) = \frac{B(z)}{A(z)} G(z) \tag{3.5}$$

From this it is clear that using this prediction model the speech production is modelled as an excitation signal, $g(n)$, filtered by a filter containing poles and zeros. There are p poles which are described by the a_i-coefficients and q zeros described by the b_j-coefficients.

An error spectrum is defined as

$$\begin{aligned} E_1(z) &= S(z) - \hat{S}(z) \\ &= S(z)A(z) - B(z)G(z) \end{aligned} \tag{3.6}$$

This corresponds to the error signal

$$e_1(n) = \sum_{i=0}^{p} a_i s(n-i) - \sum_{j=0}^{q} b_j g(n-j) \tag{3.7}$$

For the purpose of optimization of the predictor coefficients an error term is defined as

$$J_1 = \sum_{n=0}^{N-1} e_1^2(n) \tag{3.8}$$

For correct system identification of the human speech production system it is required that the recorded speech signal is undistorted in both amplitude and phase. Phase distortions may be equalized as described in appendix C.

The fixed limits of the sum have the implication of applying a rectangular window to the error signal $e_1(n)$, which corresponds to a certain range of both speech and excitation samples.

The optimal coefficients are taken as the ones that minimise J_1. The minimum is found where the derivative of J_1 with respect to all of the coefficients is zero. First the derivative with respect to a_k is found

$$\begin{aligned} \frac{\partial J_1}{\partial a_k} &= \sum_{n=0}^{N-1} \frac{\partial J_1}{\partial e_1(n)} \frac{\partial e_1(n)}{\partial a_k} = 2 \sum_{n=0}^{N-1} e_1(n) s(n-k) \\ &= 2 \sum_{n=0}^{N-1} s(n-k) \sum_{i=0}^{p} a_i s(n-i) - 2 \sum_{n=0}^{N-1} s(n-k) \sum_{j=0}^{q} b_j g(n-j) \end{aligned} \tag{3.9}$$

The derivative is set to zero and the equation is rearranged

$$\forall k \in [1;p]: \sum_{n=0}^{N-1} s(n-k) \sum_{i=0}^{p} a_i s(n-i) = \sum_{n=0}^{N-1} s(n-k) \sum_{j=0}^{q} b_j g(n-j) \Leftrightarrow \tag{3.10}$$

$$\forall k \in [1;p]: \sum_{i=0}^{p} a_i \sum_{n=0}^{N-1} s(n-k)s(n-i) = \sum_{j=0}^{q} b_j \sum_{n=0}^{N-1} s(n-k)g(n-j) \tag{3.11}$$

The same is carried out for the derivative of J_1 with respect to b_j

$$\begin{aligned} \frac{\partial J_1}{\partial b_k} &= \sum_{n=0}^{N-1} \frac{\partial J_1}{\partial e_1(n)} \frac{\partial e_1(n)}{\partial b_k} = -2 \sum_{n=0}^{N-1} e_1(n) g(n-k) \\ &= -2 \sum_{n=0}^{N-1} g(n-k) \sum_{i=0}^{p} a_i s(n-i) + 2 \sum_{n=0}^{N-1} g(n-k) \sum_{j=0}^{q} b_j g(n-j) \end{aligned} \tag{3.12}$$

$$\forall k \in [0;q]: \sum_{i=0}^{p} a_i \sum_{n=0}^{N-1} g(n-k)s(n-i) = \sum_{j=0}^{q} b_j \sum_{n=0}^{N-1} g(n-k)g(n-j) \tag{3.13}$$

The optimal a_i and b_j coefficients are found when equations (3.11) and (3.13) are solved simultaneously. This is in fact $p+q+1$ equations with $p+q+1$ unknowns which most conveniently are solved using matrix notation. Therefore the following matrix elements are defined

$$S_{ki} = \sum_{n=0}^{N-1} s(n-k)s(n-i) \qquad 0 \le k \le p \;,\; 1 \le i \le p \tag{3.14}$$

$$X_{kj} = \sum_{n=0}^{N-1} s(n-k)g(n-j) \qquad 0 \le k \le p \;,\; 0 \le j \le q \tag{3.15}$$

$$G_{kj} = \sum_{n=0}^{N-1} g(n-k)g(n-j) \qquad 0 \le k \le q \;,\; 0 \le j \le q \tag{3.16}$$

With these definitions equations (3.11) and (3.13) can be written

$$\forall k \in [1;p]: \sum_{i=0}^{p} a_i S_{ki} = \sum_{j=0}^{q} b_j X_{kj} \tag{3.17}$$

$$\forall k \in [0;q]: \sum_{i=0}^{p} a_i X_{ik} = \sum_{j=0}^{q} b_j G_{kj} \tag{3.18}$$

or (using the fact that $a_0 = 1$)

$$\forall k \in [1;p]: \sum_{i=1}^{p} a_i S_{ki} - \sum_{j=0}^{q} b_j X_{kj} = -S_{0k} \tag{3.19}$$

$$\forall k \in [0;q]: -\sum_{i=1}^{p} a_i X_{ik} + \sum_{j=0}^{q} b_j G_{kj} = X_{0k} \tag{3.20}$$

These $p + q + 1$ equations lead directly to the matrix equivalency

$$\begin{bmatrix} S_{11} & S_{12} & \dots & S_{1p} & -X_{10} & -X_{11} & \dots & -X_{1q} \\ S_{21} & S_{22} & \dots & S_{2p} & -X_{20} & -X_{21} & \dots & -X_{2q} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ S_{p1} & S_{p2} & \dots & S_{pp} & -X_{p0} & -X_{p1} & \dots & -X_{pq} \\ -X_{10} & -X_{20} & \dots & -X_{p0} & G_{00} & G_{01} & \dots & G_{0q} \\ -X_{11} & -X_{21} & \dots & -X_{p1} & G_{10} & G_{11} & \dots & G_{1q} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ -X_{1q} & -X_{2q} & \dots & -X_{pq} & G_{q0} & G_{q1} & \dots & G_{qq} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \dots \\ a_p \\ b_0 \\ b_1 \\ \dots \\ b_q \end{bmatrix} = \begin{bmatrix} -S_{01} \\ -S_{02} \\ \dots \\ -S_{0p} \\ X_{00} \\ X_{01} \\ \dots \\ X_{0q} \end{bmatrix} \tag{3.21}$$

In this set of equations the only unknowns are the coefficients $a_1 \dots a_p$ and $b_0 \dots b_q$. Since $S_{ki} = S_{ik}$ and $G_{kj} = G_{jk}$ the matrix is symmetric and therefore it can be solved efficiently with the Cholesky decomposition algorithm [Rabiner and Schafer, 1978].

3.1.2 Iterative procedure for joint optimization of glottal signal and transfer function part

The previous section describes the algorithm for calculation of the coefficients in the transfer function in the case where both the speech signal and the glottal signal are known. In speech analysis, however, only the speech signal is known. The method adopted for each speech frame in the GARMA analysis is as follows:

1. Generate a glottal signal using a glottal signal model
2. Estimate the optimal transfer function coefficients for this particular glottal signal using the algorithm described in section 3.1.1
3. Evaluate the error term J_1
4. Unless the error is sufficiently low, adjust the parameters for the glottal signal model and go to step 1.

In the original description of the GARMA analysis [Fujisaki and Ljungqvist, 1987] a glottal signal model called Fujisaki-Ljungqvist (FL) is used

(see section B.2 on page 83). The model is a 6-parameter time domain piecewise polynomial model. In this project the 4-parameter Liljencrantz-Fant (LF) model described in section B.1 on page 79 is used, primarily because it is well known and has been used in several research projects. Furthermore initial tests of the two models showed that the LF-model did not fall into as many *blind paths* with local minima during the optimization procedure as the FL-model.

3.2 WGARMA - a modified GARMA analysis

Experiments with the GARMA analysis, as described in section 3.1, on speech signals revealed a problem with correct modelling if the signal spectrum was dominated by noise in certain frequency bands. This will be demonstrated by the experiment in the following section. In section 3.2.2 on page 44 a modified GARMA algorithm, dubbed WGARMA, is introduced which allows frequency weighting of the error spectrum in order to counter this problem.

3.2.1 Disadvantageous model identification

A synthetic voiced speech signal was generated using the LF glottal signal model (section B.1) and a synthesis model with 6 complex pole pairs and one complex zero pair. The singularities and the transfer function of the synthesis model is shown in figure 3-1.

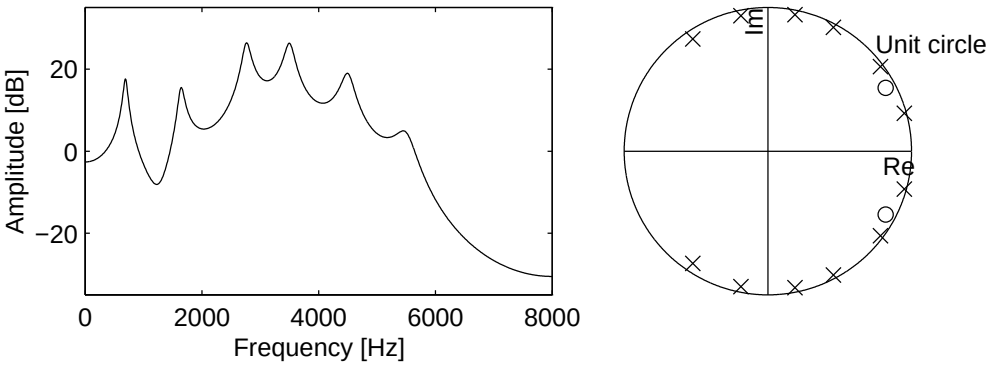


Figure 3-1: Singularities and transfer function of the synthesis model.

An FFT of the synthetic speech signal is shown in figure 3-2. As expected if a GARMA analysis is carried out on the synthetic speech with the same model order and excitation signal as was used in the generation of the synthetic signal, a model is identified which is exactly equal to the synthesis model.

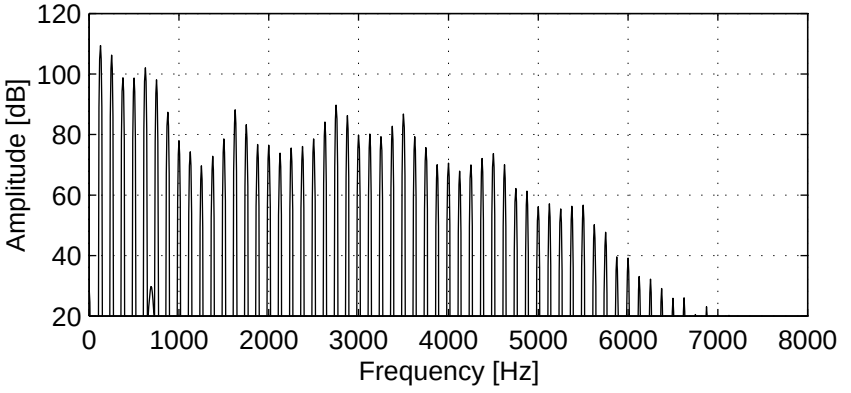


Figure 3-2: Spectrum of the synthetic speech signal (FTT of 2048-point Blackman windowed segment).

Gaussian white noise was added to the speech signal with an SNR of 40 dB. The FFT spectra of the noise and the noisy speech are shown in figures 3-3 and 3-4 respectively.

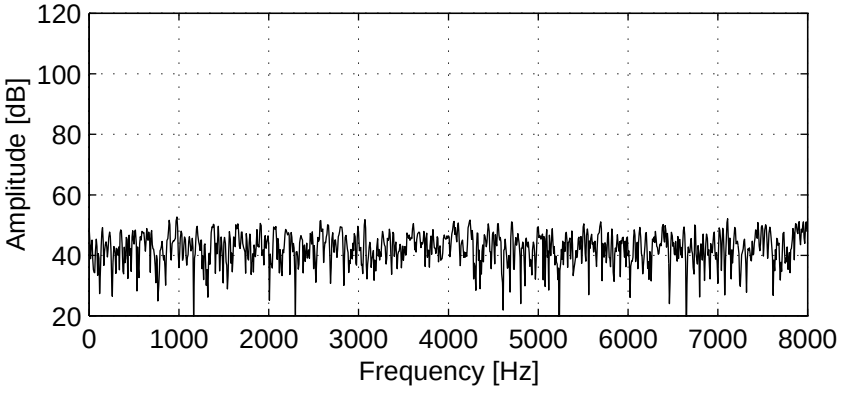


Figure 3-3: Spectrum of the noise signal (FTT of 2048-point Blackman windowed segment).

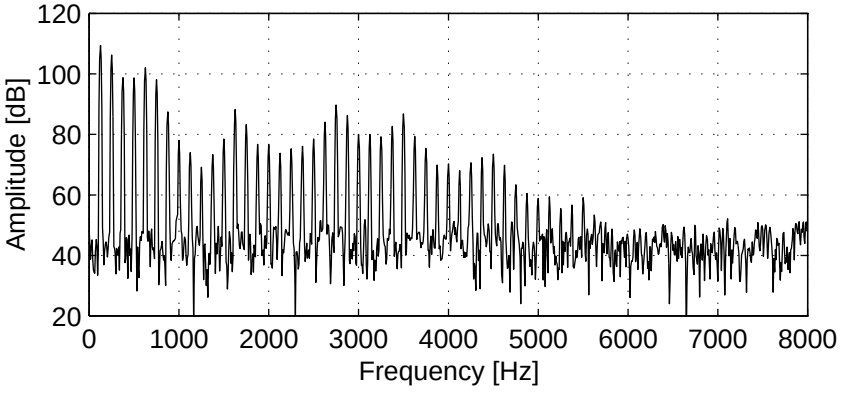


Figure 3-4: Spectrum of the noisy speech signal (FTT of 2048-point Blackman windowed segment).

A GARMA analysis was carried out on the noisy speech with the same model order and excitation signal as used in the generation of the synthetic signal. The singularities and the transfer function of the identified model is shown in figure 3-5. As is evident the model in figure 3-5 is not the same as the synthe-

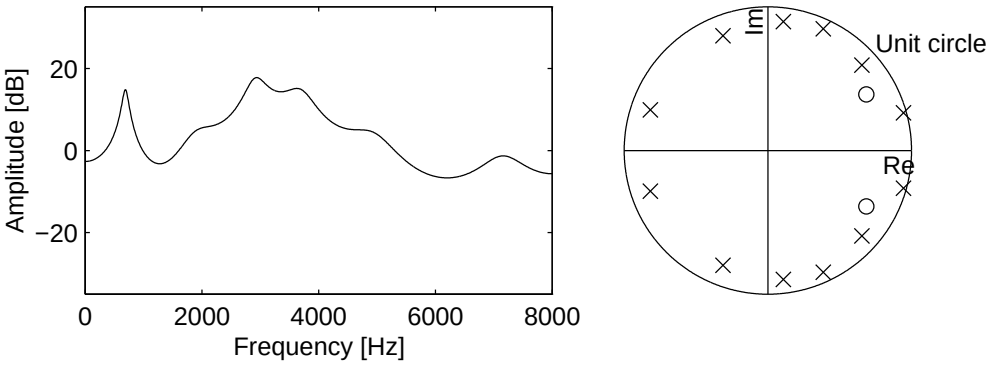


Figure 3-5: The singularities and transfer function of the model identified by the GARMA analysis. Compare to figure 3-1.

sis model in figure 3-1. This is disadvantageous in articulatory speech analysis since identification of the underlying production model is essential in recovering the shape of the vocal tract. The spectral peaks of the speech signal is more than 60 dB higher than the noise level, which is better than often experienced in recordings of real speech, but in spite of this the GARMA algorithm is not able to correctly identify the synthesis model used for the generation of the noisy speech. Apparently too much emphasis is placed by the GARMA algorithm on the noisy tail above 6000 Hz of the spectrum in figure 3-4.

In the following section the GARMA algorithm is modified to compensate for this problem.

3.2.2 An algorithm using a weighted error spectrum

In order to place more emphasis on errors at certain frequencies than others the error spectrum in equation (3.6) is modified

$$E_2(z) = W(z)E_1(z) \tag{3.22}$$

$W(z)$ is the transfer function of a weighting filter which can be interpreted as a weighting function.

Henceforth the algorithm described in this section is referred to as the *Weighted GARMA* or *WGARMA* analysis. An algorithm exists which also implements weighting of the error spectrum, but by preemphasis of both speech signal and glottal signal. This algorithm is referred to as the preemphasised GARMA algorithm. These two algorithms are compared in section 3.3 on page 49.

The error spectrum corresponds to the error signal $e'_2(n)$

$$e'_2(n) = w(n) * e_1(n) = \sum_{m=-\infty}^{\infty} w(m)\, e_1(n-m) \tag{3.23}$$

$w(n)$ is the inverse z-transform of $W(z)$. If $W(z)$ is chosen as the transfer function of an M-point FIR filter, all values of $w(n)$ except the M coefficients are zero

$$e_2(n) = \sum_{m=0}^{M-1} w(m)\, e_1(n-m)\, r_N(n-m) \tag{3.24}$$

where r_N is the same N-point rectangular window on the e_1 sequence as was implicitly applied in equation (3.8). This is to ensure that the same samples of e_1 are involved in the modified error term. e_2 is zero outside the interval $[0;\ N+M-2]$. The error term is defined similar to equation (3.8)

$$J_2 = \sum_{n=0}^{N+M-2} e_2^2(n) \tag{3.25}$$

In the following many of the same operations as in section 3.1.1 are carried out for the modified algorithm.

J_2 is differentiated with respect to a_k

$$\begin{aligned}
\frac{\partial J_2}{\partial a_k} &= \sum_{n=0}^{N+M-2} \frac{\partial J_2}{\partial e_2(n)} \frac{\partial e_2(n)}{\partial a_k} \\
&= \sum_{n=0}^{N+M-2} 2e_2(n) \sum_{m=0}^{M-1} \frac{\partial e_2(n)}{\partial e_1(n-m)} \frac{\partial e_1(n-m)}{\partial a_k} \\
&= 2 \sum_{n=0}^{N+M-2} e_2(n) \sum_{m=0}^{M-1} w(m)\, r_N(n-m)\, s(n-m-k) \qquad (3.26) \\
&= 2 \sum_{n=0}^{N+M-2} \left(\sum_{r=0}^{M-1} w(r)\, r_N(n-r) \left(\sum_{i=0}^{p} a_i s(n-r-i) - \sum_{j=0}^{q} b_j g(n-r-j) \right) \right) \\
&\qquad \sum_{m=0}^{M-1} w(m)\, r_N(n-m)\, s(n-m-k)
\end{aligned}$$

Setting the derivatives equal to zero to find the minimum and arranging the a_i and b_j elements on either side of the equation

$\forall k \in [1;p]:$

$$\sum_{i=0}^{p} a_i \sum_{n=0}^{N+M-2} \left(\sum_{m=0}^{M-1} s(n-m-k)w(m)r_N(n-m) \right) \sum_{r=0}^{M-1} s(n-r-i)w(r)r_N(n-r) \tag{3.27}$$
$$= \sum_{j=0}^{q} b_j \sum_{n=0}^{N+M-2} \left(\sum_{m=0}^{M-1} s(n-m-k)w(m)r_N(n-m) \right) \sum_{r=0}^{M-1} g(n-r-j)w(r)r_N(n-$$

The same operation is carried out with respect to b_k

$$\frac{\partial J_2}{\partial b_k} = \sum_{n=0}^{N+M-2} \frac{\partial J_2}{\partial e_2(n)} \frac{\partial e_2(n)}{\partial b_k}$$
$$= \sum_{n=0}^{N+M-2} 2e_2(n) \sum_{m=0}^{M-1} \frac{\partial e_2(n)}{\partial e_1(n-m)} \frac{\partial e_1(n-m)}{\partial b_k}$$
$$= -2 \sum_{n=0}^{N+M-2} e_2(n) \sum_{m=0}^{M-1} w(m)r_N(n-m)g(n-m-k) \tag{3.28}$$
$$= -2 \sum_{n=0}^{N+M-2} \left(\sum_{r=0}^{M-1} w(r)r_N(n-r) \left(\sum_{i=0}^{p} a_i s(n-r-i) - \sum_{j=0}^{q} b_j g(n-r-j) \right) \right)$$
$$\sum_{m=0}^{M-1} w(m)r_N(n-m)g(n-m-k)$$

Again the derivative is set to zero and the a_i and b_j elements are isolated on either side

$\forall k \in [0;q]:$

$$\sum_{i=0}^{p} a_i \sum_{n=0}^{N+M-2} \left(\sum_{m=0}^{M-1} g(n-m-k)w(m)r_N(n-m) \right) \sum_{r=0}^{M-1} s(n-r-i)w(r)r_N(n-r) \tag{3.29}$$
$$= \sum_{j=0}^{q} b_j \sum_{n=0}^{N+M-2} \left(\sum_{m=0}^{M-1} g(n-m-k)w(m)r_N(n-m) \right) \sum_{r=0}^{M-1} g(n-r-j)w(r)r_N(n-$$

Equations (3.27) and (3.29) are similar to equations (3.11) and (3.13) respectively. Analogous with equations (3.14)-(3.16) the following is defined

$$\tilde{S}_{ki} = \sum_{n=0}^{N+M-2} \left(\sum_{m=0}^{M-1} s(n-m-k)w(m)r_N(n-m) \right) \sum_{r=0}^{M-1} s(n-r-i)w(r)r_N(n-r) \quad (3.30)$$

$$\tilde{X}_{kj} = \sum_{n=0}^{N+M-2} \left(\sum_{m=0}^{M-1} s(n-m-k)w(m)r_N(n-m) \right) \sum_{r=0}^{M-1} g(n-r-j)w(r)r_N(n-r) \quad (3.31)$$

$$\tilde{G}_{kj} = \sum_{n=0}^{N+M-2} \left(\sum_{m=0}^{M-1} g(n-m-k)w(m)r_N(n-m) \right) \sum_{r=0}^{M-1} g(n-r-j)w(r)r_N(n-r) \quad (3.32)$$

To facilitate the calculation of equations (3.30)-(3.32) two definitions are made

$$\sigma_{kn} = \sum_{m=0}^{M-1} s(n-m-k)w(m)r_N(n-m) \qquad 0 \le k \le p, \quad 0 \le n \le N+M-2 \quad (3.33)$$

$$\gamma_{jn} = \sum_{r=0}^{M-1} g(n-r-j)w(r)r_N(n-r) \qquad 0 \le k \le q, \quad 0 \le n \le N+M-2 \quad (3.34)$$

Which simplifies equations (3.30)-(3.32)

$$\tilde{S}_{ki} = \sum_{n=0}^{N+M-2} \sigma_{kn}\sigma_{in} \qquad 0 \le k \le p, \quad 1 \le i \le p \quad (3.35)$$

$$\tilde{X}_{kj} = \sum_{n=0}^{N+M-2} \sigma_{kn}\gamma_{jn} \qquad 0 \le k \le p, \quad 0 \le j \le q \quad (3.36)$$

$$\tilde{G}_{kj} = \sum_{n=0}^{N+M-2} \gamma_{kn}\gamma_{jn} \qquad 0 \le k \le q, \quad 0 \le j \le q \quad (3.37)$$

Exactly as in equation (3.21) the $p+q+1$ equations of equations (3.27) and (3.29) can be solved conveniently with matrix notation

$$\begin{bmatrix} \tilde{S}_{11} & \tilde{S}_{12} & \dots & \tilde{S}_{1p} & -\tilde{X}_{10} & -\tilde{X}_{11} & \dots & -\tilde{X}_{1q} \\ \tilde{S}_{21} & \tilde{S}_{22} & \dots & \tilde{S}_{2p} & -\tilde{X}_{20} & -\tilde{X}_{21} & \dots & -\tilde{X}_{2q} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \tilde{S}_{p1} & \tilde{S}_{p2} & \dots & \tilde{S}_{pp} & -\tilde{X}_{p0} & -\tilde{X}_{p1} & \dots & -\tilde{X}_{pq} \\ -\tilde{X}_{10} & -\tilde{X}_{20} & \dots & -\tilde{X}_{p0} & \tilde{G}_{00} & \tilde{G}_{01} & \dots & \tilde{G}_{0q} \\ -\tilde{X}_{11} & -\tilde{X}_{21} & \dots & -\tilde{X}_{p1} & \tilde{G}_{10} & \tilde{G}_{11} & \dots & \tilde{G}_{1q} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ -\tilde{X}_{1q} & -\tilde{X}_{2q} & \dots & -\tilde{X}_{pq} & \tilde{G}_{q0} & \tilde{G}_{q1} & \dots & \tilde{G}_{qq} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \dots \\ a_p \\ b_0 \\ b_1 \\ \dots \\ b_q \end{bmatrix} = \begin{bmatrix} -\tilde{S}_{01} \\ -\tilde{S}_{02} \\ \dots \\ -\tilde{S}_{0p} \\ \tilde{X}_{00} \\ \tilde{X}_{01} \\ \dots \\ \tilde{X}_{0q} \end{bmatrix} \quad (3.38)$$

By inspection of the elements it is evident that also this matrix is symmetric and consequently the equation can be solved efficiently with the Cholesky algorithm.

In continuation of the experiment in section 3.2.1 a WGARMA analysis was carried out on the noisy speech signal shown in figure 3-4 (page 43). As an initial test the weighting filter was more or less randomly chosen as a 41-tap linear phase FIR filter with the transfer function shown in figure 3-6. The filter has

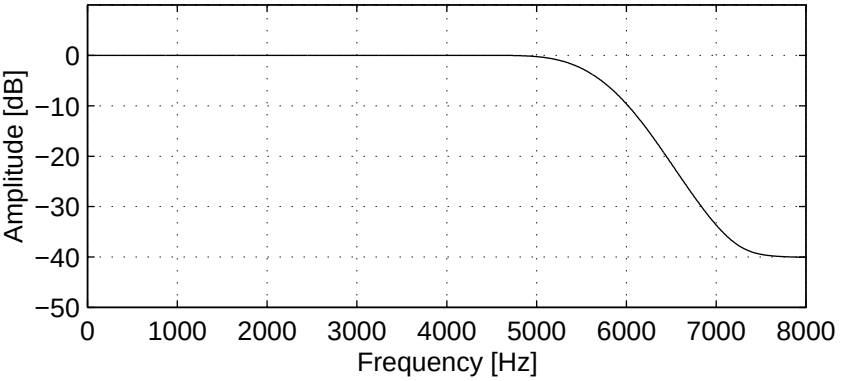


Figure 3-6: The transfer function of the weighting filter in the WGARMA analysis.

unity gain from DC to 5 kHz and an attenuation of around 40 dB above 7.5 kHz. When the WGARMA analysis was carried out the singularities identified where identical to the singularities of the model used for the generation of the synthetic speech signal. This can be seen if figure 3-7 is compared to figure 3-1.

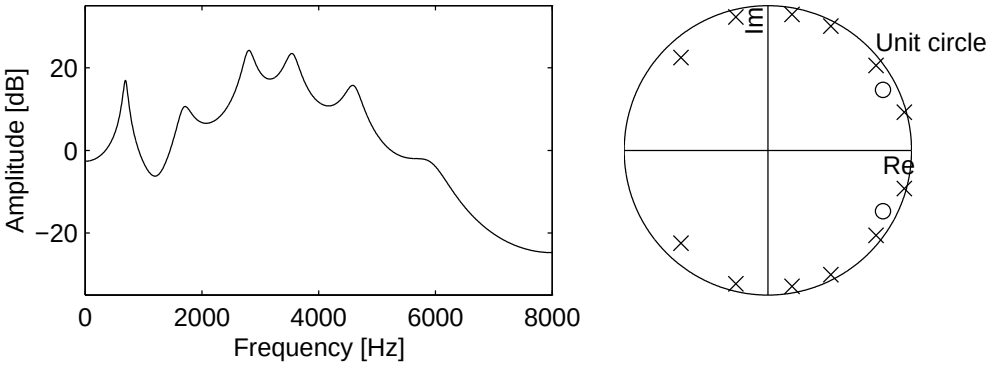


Figure 3-7: The singularities and transfer function of the model identified by the WGARMA analysis.

As shown the WGARMA algorithm is able to weigh certain frequencies higher than others when performing the system identification. By adjusting the filter $W(z)$ this weighting function is very flexible.

3.3 Comparison of WGARMA implementations

An alternative to the WGARMA algorithm exists which also implements frequency weighting of the error spectrum. In the following this algorithm will be dubbed the *preemphasised GARMA* algorithm. In essence the weighting is obtained by preemphasising both the speech signal and the excitation signal using the weighting filter $W(z)$ and subsequently a normal GARMA analysis is performed. In this section the two alternative algorithms are compared in terms of system identification capabilities of steady state signals and signals with rapidly changing underlying synthesis models as well as in terms of computational requirements.

3.3.1 System level comparison

To clarify the differences between the GARMA, WGARMA and preemphasised GARMA analyses they are viewed as block diagrams in the frequency domain. First the GARMA analysis is shown in the context of the underlying speech synthesis model $\tilde{B}(z)/\tilde{A}(z)$ and excitation signal $\tilde{G}(z)$, see figure 3-8.

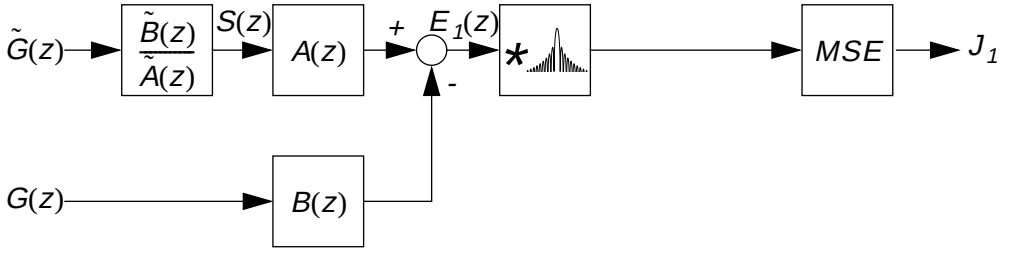


Figure 3-8: Frequency domain block diagram of a GARMA analysis.

The synthesis model and excitation signal are underlying in the sense that they are unknown in the case of real speech signals. In the algorithm the parameters of the $A(z)$ and $B(z)$ functions are optimised to achieve minimum variance of the signal entering the MSE block (Minimum Squared Error). If $G(z) \approx \tilde{G}(z)$ then this is obtained if $A(z)$ approaches $\tilde{A}(z)$ and $B(z)$ approaches $\tilde{B}(z)$ which corresponds to correct system identification. The window function, which is implicit in equation (3.8), corresponds in the frequency domain to a convolution with the transfer function of the window as shown.

The corresponding block diagram for the WGARMA algorithm is given in figure 3-9. Compared to the GARMA analysis the only difference is the weighting filter inserted before the MSE block.

In the preemphasised GARMA algorithm the speech signal and the excitation signal are preemphasised with the $W(z)$ filter before they are analysed by a GARMA algorithm, see figure 3-10a. In figure 3-10b an equivalent system is

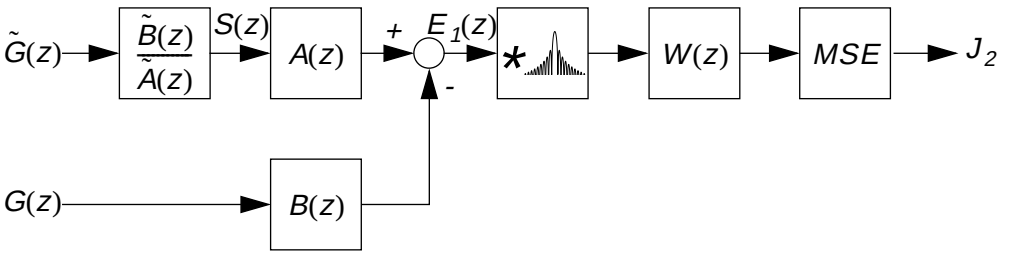


Figure 3-9: Frequency domain block diagram of a WGARMA analysis.

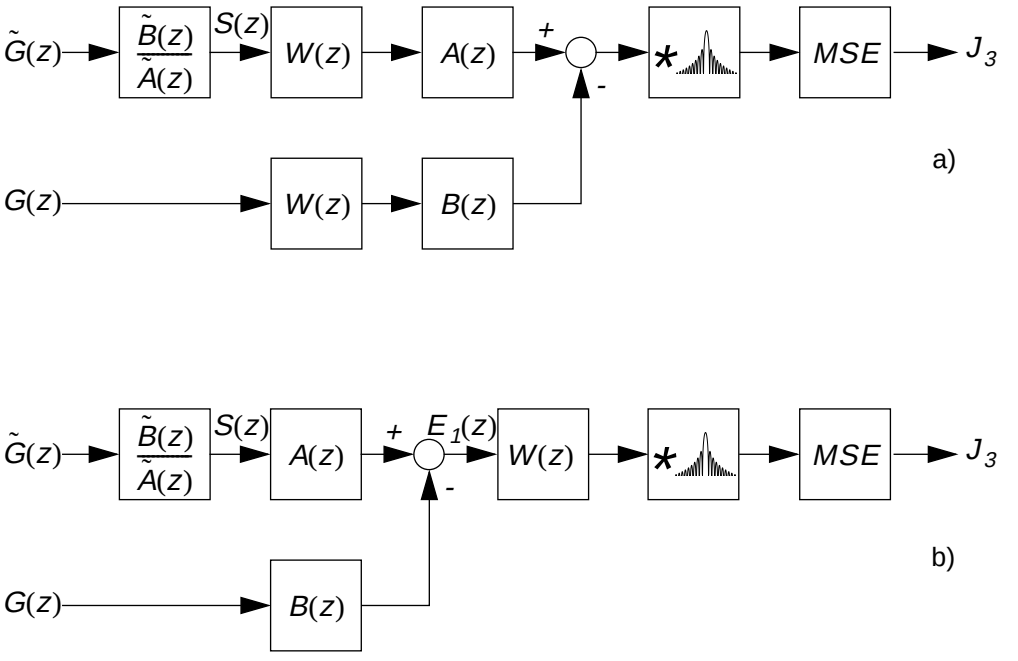


Figure 3-10: a) Frequency domain block diagram of a preemphasised GARMA analysis. b) Equivalent system with preemphasis filter moved.

shown where the preemphasis filters have been moved. This is equivalent because the $A(z)$, $B(z)$ and $W(z)$ blocks are linear. If figure 3-9 and 3-10b are compared it is clear that the WGARMA and the preemphasised GARMA algorithms differ solely in the order of the blocks: in the WGARMA algorithm the window is applied before the weighting filter and in the preemphasised GARMA algorithm the order of the two blocks is reversed.

3.3.2 Computational loads

As described in section 3.1.2 on page 41 the GARMA algorithm and the just described variants are iterative procedures. For each frame of speech signal several excitation signals are tested by carrying out the optimization and evaluating the corresponding error term. Typically the number of tested excitation

signals for each speech frame is several hundreds or thousands. When comparing the computational loads of the WGARMA and preemphasised GARMA algorithms, the crucial factor therefore is the computational load of a single iteration in which the excitation signal is changed but the speech signal is kept unchanged.

Although the expressions in the derivation of the WGARMA algorithm in section 3.2.2 are more complex than the GARMA counterparts in section 3.1.1, the final equations (3.35)-(3.38) are quite similar to the corresponding GARMA equations (3.14)-(3.16) and (3.21). In the WGARMA algorithm the γ_{jn}'s of equation (3.34) must be computed when the excitation signal is changed. This is not the case for the preemphasised GARMA algorithm, which on the other hand requires that the excitation signal is filtered by the weighting filter.

The computational loads of both algorithms depend on programming details as well as hardware capabilities neither of which are considered here, but table 3-1 is an estimate of the number of multiply-add operations involved in the two algorithms compared.

	Preemphasised GARMA		WGARMA	
Preemphasis	$M(N+q+M)$	3843	-	-
γ	-	-	$M(N+q+M)$	3843
X	$(p+1)(q+1)N$	6240	$(p+1)(q+1)(N+M-1)$	7020
G	$\frac{1}{2}(q^2+q)N$	480	$\frac{1}{2}(q^2+q)(N+M-1)$	540
Cholesky	$\frac{(p+q+1)^3}{3}+(p+q+1)^2-\frac{4}{3}(p+q+1)$			1330
Glottal signal	$100+10(N+q+1)$			1730
Total	13623		14463	

Table 3-1: Estimate of required number of multiply-add operations for one iteration for the preemphasised GARMA and the WGARMA algorithms. Actual numbers are in the case M=21, N=160, p=12 and q=2.

Generally the WGARMA algorithm constitutes a higher computational load than the preemphasised GARMA algorithm, but for typical values of N, M, p and q the difference is negligible. In the case shown in table 3-1 the load is 6% higher for the WGARMA algorithm compared to the preemphasised GARMA algorithm.

3.3.3 System identification capabilities

As expected the WGARMA and preemphasised GARMA algorithms generally perform similarly as system identification algorithms. This is particularly pronounced if M is small compared to N. However in some cases the WGARMA algorithm performs superiorly in tracking rapid changes in the underlying production model. Conversely this is most striking when M is not small compared to N. The explanation can be found when comparing figure 3-10b with figure 3-9. The two systems are identical except that in the preemphasised GARMA algorithm the preemphasis filter is applied before the window and in the WGARMA algorithm it is applied after the window. In the time domain this corresponds to a convolution with the impulse response of the weighting filter, which has the length M. Because of this convolution the preemphasised GARMA algorithm spreads very localized properties of the signal over a wider segment in time before the error signal enters the MSE block, which determines which model parameters are identified. In the WGARMA algorithm only the error samples selected by the window contribute to the error measure and therefore more local effects can be identified.

The difference in model identification performance is demonstrated in the following example.

A synthetic speech signal was generated the same way as in section 3.2.1, the coefficients in the synthesis model were changed instantaneously corresponding to an immediate decrease of all formant frequencies and zeros by 500 Hz. Figure 3-11 shows the transfer function of the synthesis model before and after the transition.

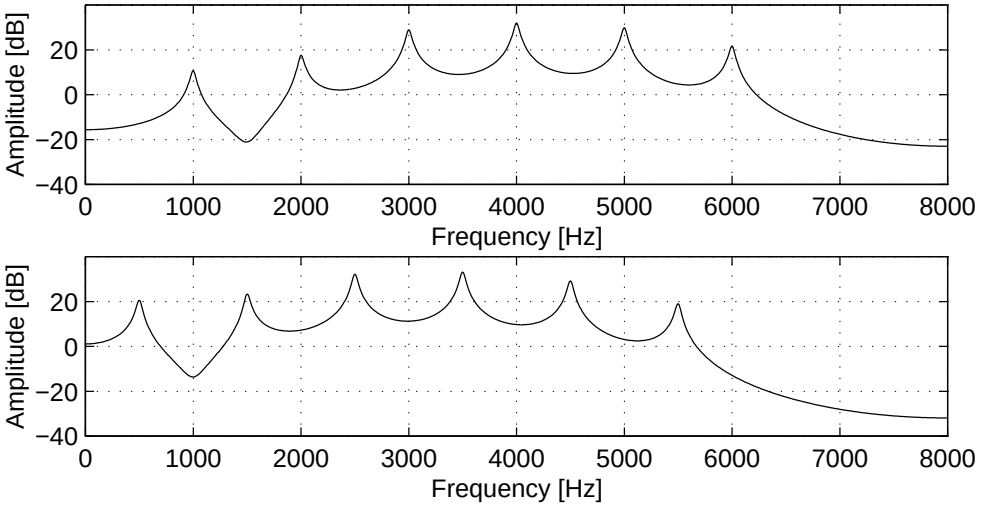


Figure 3-11: Transfer function of the synthesis model before (top) and after (bottom) the transition during the generation of the synthetic speech signal.

A spectrogram-like presentation of the transfer function of the synthesis model and the evolution in time is given in figure 3-12. The darker areas in the figure correspond to the spectral peaks in figure 3-11.

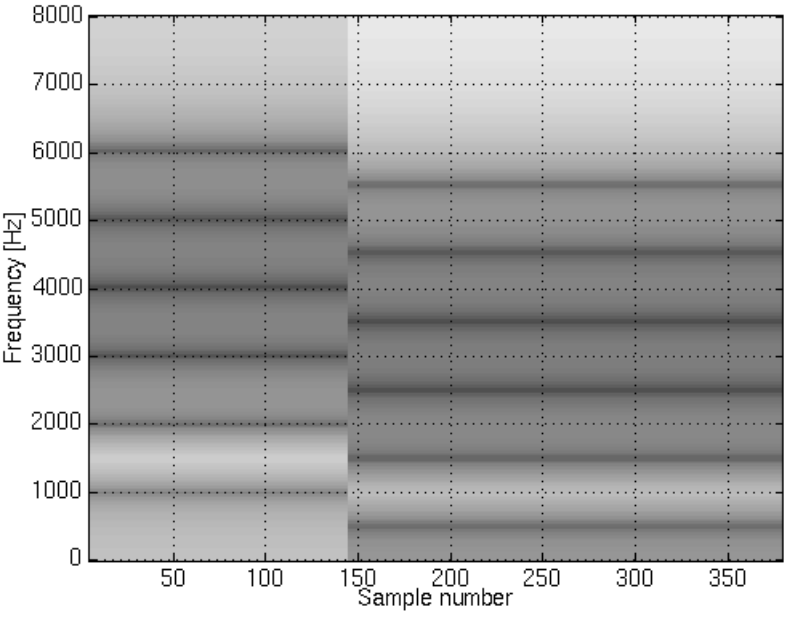
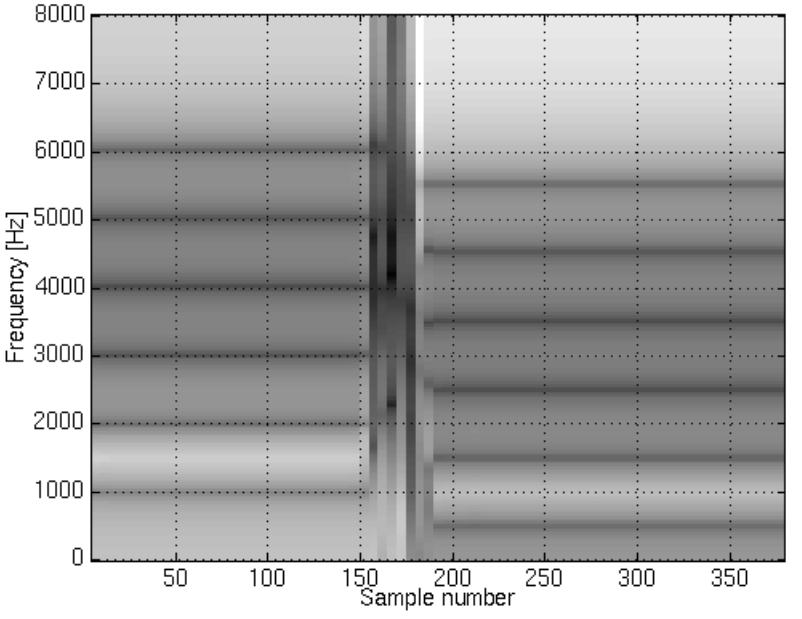


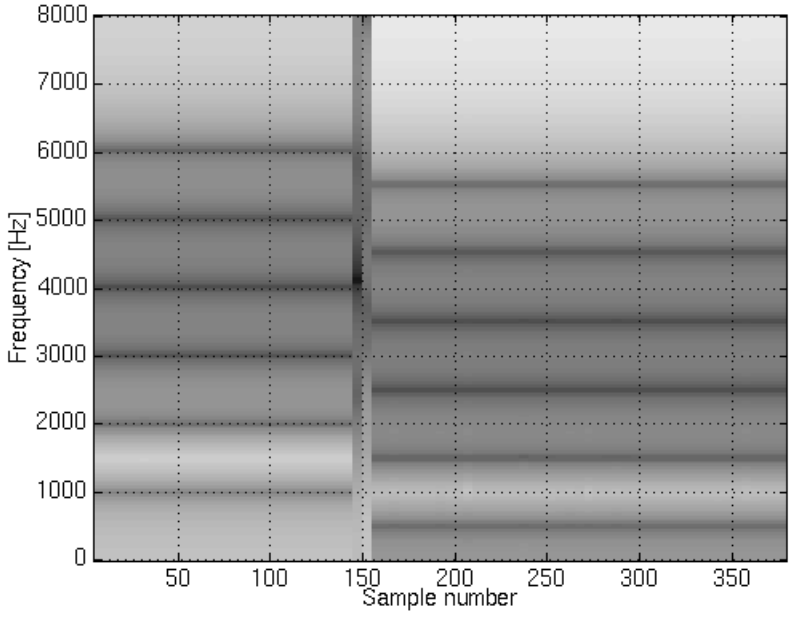
Figure 3-12: Spectrogram-like representation of the transfer function of the synthesis model used for the generation of the synthetic speech signal.

An emphasised GARMA analysis and a WGARMA analysis were carried out every 5 samples on the synthetic speech signal using the original excitation signal. The same 41-tap weighting filter as used earlier was employed (figure 3-6) and the analysis window was narrowed to 15 samples in order to exaggerate the difference between the two algorithms. The results are shown in figure 3-13.

From figure 3-13 it is clear that under certain circumstances the WGARMA algorithm can follow faster changes in the system on which the system identification is being performed. The WGARMA algorithm is limited in this respect only by the window length. In figure 3-13b only the two frames in which the model change appears inside are affected. Figure 3-13a, however, shows that several frames are affected by the convolution with the weighting filter impulse response.



a)



b)

Figure 3-13: Time evolution of the transfer function of the model identified by a) the preemphasised GARMA algorithm and b) the WGARMA algorithm.

3.4 Summary

In this chapter signal analysis algorithms corresponding to the speech production model described in chapter 2 are discussed. The system level function

of the signal analysis algorithm is to analyse a frame of speech signal and obtain one or more sets of poles and zeros.

The WGARMA algorithm is proposed, which is a GARMA algorithm which facilitates frequency weighting of the error signal. The WGARMA algorithm is shown to have certain advantages compared to a GARMA analysis operating on preemphasised signals.

The signal analysis algorithm may result in parameters that are suboptimal with respect to their correspondence with the best possible match of the actual vocal tract shape of the speaker during production. This may be caused by a number of factors:

- Acoustic noise in the speech signal
- Imperfections of the speech production model
- Imperfections in the glottal signal model
- Local minima in the search for optimal glottal parameters

Because the optimal parameters are not always the ones that correspond to the minimal error term (equation (3.25)) the result of the signal analysis algorithm may be given as a number of conjectured sets of poles and zeros. The selection of these sets among the many sets tested in the search for glottal parameters is a subject for further investigation.

4 Application in articulatory speech analysis

In this chapter the integration and application of the two components described in the preceding chapters is discussed - the speech production model and the signal analysis algorithm. Furthermore future work related to this project is suggested.

4.1 Application of proposed components

As mentioned on page 11 this report is not an attempt to solve the problem of articulatory speech analysis as such. The main focus of the report is in the area of improvements and investigations of some of the fundamental components in speech production modelling and signal analysis algorithms for articulatory speech analysis. Some preliminary tests were carried out on the recorded speech signals using prototype C++-language implementations of the algorithms. The next section will attempt to show that the combination of the described components to form a complete articulatory speech analysis system is a complex task. In section 4.1.2 some possible approaches to the task are proposed.

4.1.1 Problem outline

In this section some elements and aspects of the integration of the speech production model and the signal analysis algorithm into a complete articulatory speech analysis system are described.

To summarize the characteristics on a system level of the two components treated in this report:

- The speech production model, given the vocal tract shape, is able to determine the corresponding poles and zeros in the transfer function.
- The signal analysis algorithm analyses a frame of speech signal and determines one or more sets of conjectured poles and zeros of the speech production model. Each set of poles and zeros is accompanied by a parameterized glottal signal.

As mentioned in chapter 1 the goal of articulatory speech analysis is to determine the vocal tract shape for each speech frame. This is not directly possible using the two components since the speech production model is not invertible. In other words the vocal tract shape is not directly obtainable from the poles and zeros resulting from the signal analysis algorithm. These relations are attempted illustrated in figure 4-1.

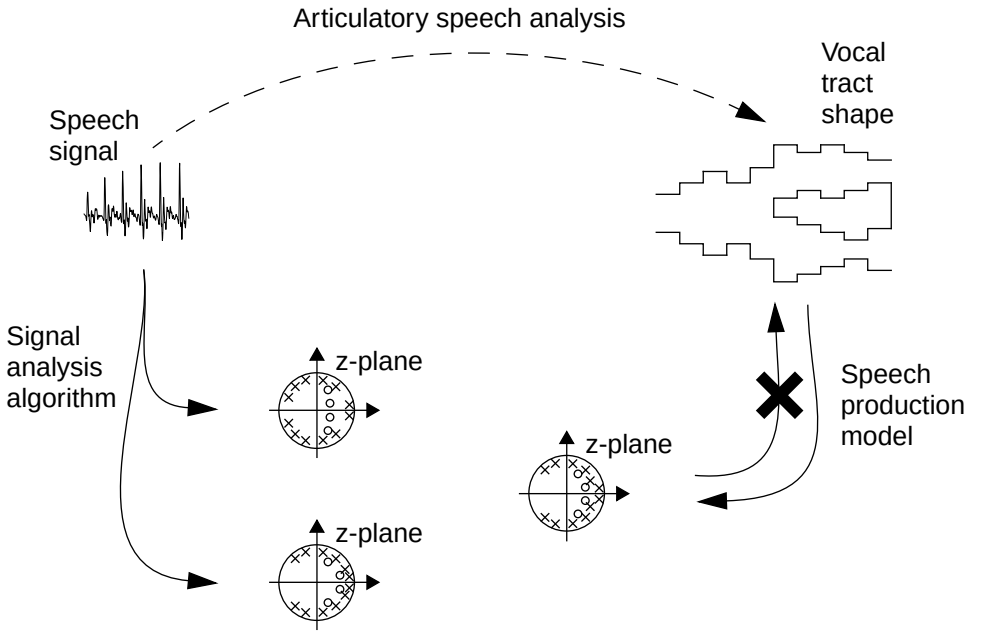


Figure 4-1: System level view of the signal analysis algorithm and the speech production model compared to articulatory speech analysis. The crossed out arrow indicates that the vocal tract shape can not be determined directly from the poles and zeros.

Since direct inversion is not possible methods must be found to

1. Optimise and select the parameters for the glottal signal model to produce one or more conjectured sets of poles and zeros.

2. Select a vocal tract shape that matches one of these sets.

Ad 1: Because of imperfections in the speech production model and glottal signal model, local minima in the iterative search or acoustic noise it is not necessarily the parameters that yield the lowest error term (equation (3.25) on page 45) that are optimal in the sense that they result in a vocal tract shape closest to the one involved in the actual production of the speech.

Ad 2: Multiple vocal tract shapes may correspond to a specific set of poles and zeros within a limited margin of error. This is the direct manifestation of the non-uniqueness in speech production described on page 10.

Both of these selections involve several parameters and may not be obvious, since the correct choice depends on many complex factors, such as which vocal tract shape is closest to the original.

4.1.2 Proposed approaches

In this section a number of approaches to the solution of the problems outlined in the previous section are proposed. Many of the approaches involve application of other knowledge sources to constrain the search for a vocal tract shape.

An important knowledge source, that most likely is essential in articulatory speech analysis, is the knowledge of how the vocal tract shape is constrained. This may be expressed simply as limits on the cross-sectional areas of each tube section in the speech production model. Preferably, however, the constraints are expressed in the form of an articulatory model as described in section 1.3.1 on page 6. Apart from imposing constraints on the vocal tract shape, an articulatory model also reduces the dimension of the search space because the number of parameters of the articulatory model is significantly smaller than the number of tube sections in the speech production model. An obvious constraint on the vocal tract shape is the nasal cavity which is fixed for each speaker.

Another important constraint is the property of continuity in time of the vocal tract shape. The speech signal itself or any of the parameter sets that are closely related to the acoustic signal, such as LPC or cepstral coefficients, exhibit quasistationarity in many speech sounds. Certain sound patterns, however, result in sudden changes in the acoustic domain. Because of the physical properties of the speech production organs, such as mass, force, and friction, the vocal tract shape is considerably more constrained in the possible change

from one frame to the next. A number of studies in articulatory phonetic have revealed limits of movement for the various articulators e.g. the root of the tongue is limited to slower movements than the tip of the tongue or the lips. These properties are perhaps also most efficiently modelled with the help of an articulatory model.

The glottal signal parameters too have constraints both in value and time. Experience with the signal analysis algorithm will limit the search space for the glottal parameters. More importantly the glottal signal is very periodic and changes very slowly compared to the articulatory movements. As shown in appendix C the speech recordings may be accompanied by synchronized recordings of the signal from an accelerometer mounted on the speakers larynx. An analysis of this signal will yield the fundamental frequency and an estimate of the time alignment of the glottal signal.

As a starting point for the search for a vocal tract shape that matches a given set of poles and zeros a codebook may be generated by sampling randomly among the possible vocal tract shapes and storing the corresponding calculated poles and zeros in a table. Using an appropriate distance measure, which also is a subject for investigation, the stored pole and zero sets that are closest to the one given are looked up in the codebook. The corresponding vocal tract shapes can be used as starting points in the search process. The same technique could be used initially to determine the shape of the nasal cavity for a given speaker. Once enough evidence for a certain shape has been found, the shape of the nasal cavity can be held fixed and only the coupling to the rest of the vocal tract must be optimised.

An analysis may be carried out in parallel using an all-pole speech production model. For non-nasalized speech sounds this analysis will be as accurate as the pole-zero analysis described in this report. As mentioned on page 10 this type of analysis has the advantage that the speech production model is invertible, which means that the vocal tract shape can be calculated directly from the poles.

Since the articulatory speech analysis typically is carried out off line, the analysis may start at the signal parts with little ambiguity and apply the time-continuity constraints both forwards and backwards in time, in order to supply stronger constraints on the optimization in the more problematic segments of the signal.

4.2 Long term perspectives

Many of the proposed approaches in the previous section must be examined in order to integrate the speech production model and the signal analysis algorithm into an articulatory speech analysis system.

Among the further future work topics could be the combination of articulatory speech analysis with some of the other techniques described in section 1.4. Perhaps most promising in this respect is magnetic resonance imaging (MRI) described in section 1.4.2. The two types of analyses supplement each other well. MRI-films can yield 3-dimensional image sequences of the vocal tract shape, but the images are very noise contaminated and the time resolution is poor. These data, however, could prove as valuable starting points for an articulatory speech analysis of the acoustic signal, which in this case would act as a refinement of the MRI analysis.

5 Conclusion

The work documented in this report is in the area of articulatory analysis of speech signals. The long term objective of this research area is to be able to, by means of analysis of a speech signal, to determine the details of how the speech signal was produced and most importantly the vocal tract shape as a function of time. Some of the elements of such an articulatory analysis are proposed in this report.

A speech production model is established which is enhanced compared to the model corresponding to LPC analysis. The enhanced model includes a model of the nasal cavity thereby making articulatory correct analysis of nasalized speech sounds possible. Based on the same assumptions as the model corresponding to LPC analysis, a chain of tube sections modelling the nasal cavity is added in a Y-junction and the equivalent time domain system is determined. None of the three resulting chains of tube sections can be regarded as independent in this respect. The transfer function of the model is determined, but although this is an important result, the expression is so complex that the applications are limited. Nevertheless, by means of programs for symbolic mathematics, the order of the transfer function is determined. This result in turn

allows the determination of the transfer function on the rational function form by system identification on the time domain system. Hence on a system level, by using the speech production model and the mentioned techniques, it is possible to determine the poles and zeros of the transfer function, given the cross-sectional areas of all the tube sections in the model.

A signal analysis algorithm called WGARMA is developed, which identifies the poles and zeros of the speech production model corresponding to a segment of speech signal. The algorithm optimises the parameters of a glottal signal model such that the frequency weighted prediction error of the model becomes minimum. The WGARMA algorithm is shown be advantageous compared to a preemphasised GARMA analysis.

A database of speech signals is recorded in an anechoic room. Techniques for the measurement of the recording equipment transfer function are developed together with methods to equalize the recordings in order to obtain linear amplitude and phase characteristics.

The described elements are proposed as basis for further work in the area of articulatory speech analysis.

Among other things strategies must be found to optimise the parameters for the glottal signal model. Furthermore the mapping from singularities found by the signal analysis algorithm to vocal tract shapes must be investigated. The mapping is not guaranteed to be one-to-one, but the choice among multiple shapes can be aided by various constraints on the articulation process.

It is suggested that articulatory analysis of speech signals could be supplemented by other forms of analyses such as glottal signal analysis or MRI in order to obtain good starting points for the optimization. In this case the speech analysis algorithm would act as a refining process in both time and shape.

A further development of the speech production model would be to incorporate losses due to thermal and viscous properties among others in the vocal tract. Acoustic losses are likely to be relatively important in the nasal cavity compared to the pharynx and oral cavity.

A Order of the transfer function

In this appendix the number of poles and zeros will be determined for the transfer function of the speech production model derived in section 2.4. The appendix should be read in conjunction with section 2.5 starting on page 31.

As will be evident in the following pages the joint complexity of the equations expressing the transfer function is such that manual rewriting to determine the order has not been possible. Therefore programs for symbolic mathematics have been applied for this purpose. The commercial programs **MapleV**[1] and **Mathematica**[2] were tested and yielded the same results.

The following pages reproduce a **MapleV**-session (on the right pages) and the corresponding explanation (on the left pages). Input to and output from **MapleV** is represented with dissimilar typefaces:

```
> y:=x^2;
```
← Input (source line from program)

$$y := x^2$$ ← Output from Maple

1. **MapleV© release 3** from the Symbolic Computation Group at the University of Waterloo, [Char et al., 1992], [Char et al., 1991a], [Char et al., 1991b].

2. **Mathematica©** from Wolfram Research, Inc. [Wolfram, 1991].

A.1 MapleV-program

The text file named `transfer_function_order` contains 22 lines of **MapleV** statements. The program is read into maple with the statement:

```
> read transfer_function_order;
```

In this session the input statements from the program are echoed by **MapleV** to facilitate readability.

The `with` statement reads in the library for linear algebra, which enables matrix algebra.

```
> with(linalg, matrix):
```

The next program line assigns the matrix expression on the right to the name `QQn`, which denotes the product of $\boldsymbol{Q}$-matrices. This is equivalent to equation (2.70) on page 32, where it is mentioned that N_i (`Ni` in this appendix) is a polynomial in z^{-1} of the order $M_N - 2$.

```
> QQn := array([[N1, N2/z],[N3, N4/z]]);
```

Similarly for the oral tract and the pharynx:

```
> QQo := array([[O1, O2/z],[O3, O4/z]]);
> QQp := array([[P1, P2/z],[P3, P4/z]]);
```

The following four lines implement R_N and R_O as equations (2.50)-(2.51):

```
> tmp:=evalm(QQn &* [[ 1 ], [ -mu[n1] ]] ):
> Rn:= tmp[1,1]/tmp[2,1] * z;
> tmp:=evalm(QQo &* [[ 1 ], [ -mu[o1] ]]):
> Ro:= tmp[1,1]/tmp[2,1] * z;
```

The `evalm` function with the `&*` operator implements matrix multiplication. Notice in **MapleV**'s output that prior definitions are used in subsequent expressions.

H_A is defined from equation (2.61):

```
> HA:=evalm(2/(1+mu[pg])*[[1,-mu[pg]*z^(-1)]]*z^(Mp/2)&*
QQp&*[[1],[Rp]])[1,1];
```

R_P is defined from equation (2.58):

```
> Rp:=mu[pj]+(1+mu[pj])*(HB+HD);
```

H_E, H_D and H_C are defined from equations (2.62), (2.54) and (2.63) respectively:

```
> HE:=evalm([[0,1]]*z^(Mn/2-1)&*QQn&*[[1],[-mu[n1]]]*1/
(1+mu[n1]))[1,1];
> HD:=(1+mu[nj])/(Rn-mu[nj])*(1+HB);
> HC:=evalm([[0,1]]*z^(Mo/2-1)&*QQo&*[[1],[-mu[o1]]]*1/
(1+mu[o1]))[1,1];
```

```
> read transfer_function_order;
> with(linalg, matrix):
> QQn := array([[N1, N2/z],[N3, N4/z]]);
```

$$\mathrm{QQn} := \begin{bmatrix} \mathrm{N1} & \dfrac{\mathrm{N2}}{z} \\ \mathrm{N3} & \dfrac{\mathrm{N4}}{z} \end{bmatrix}$$

```
> QQo := array([[O1, O2/z],[O3, O4/z]]);
```

$$\mathrm{QQo} := \begin{bmatrix} \mathrm{O1} & \dfrac{\mathrm{O2}}{z} \\ \mathrm{O3} & \dfrac{\mathrm{O4}}{z} \end{bmatrix}$$

```
> QQp := array([[P1, P2/z],[P3, P4/z]]);
```

$$\mathrm{QQp} := \begin{bmatrix} \mathrm{P1} & \dfrac{\mathrm{P2}}{z} \\ \mathrm{P3} & \dfrac{\mathrm{P4}}{z} \end{bmatrix}$$

```
> tmp:=evalm(QQn &* [[ 1 ], [ -mu[n1] ]] ):
> Rn:= tmp[1,1]/tmp[2,1] * z;
```

$$\mathrm{Rn} := \frac{(\mathrm{N1}\,z - \mathrm{N2}\,\mu_{\mathrm{n1}})\,z}{\mathrm{N3}\,z - \mathrm{N4}\,\mu_{\mathrm{n1}}}$$

```
> tmp:=evalm(QQo &* [[ 1 ], [ -mu[o1] ]]):
> Ro:= tmp[1,1]/tmp[2,1] * z;
```

$$\mathrm{Ro} := \frac{(\mathrm{O1}\,z - \mathrm{O2}\,\mu_{\mathrm{o1}})\,z}{\mathrm{O3}\,z - \mathrm{O4}\,\mu_{\mathrm{o1}}}$$

```
> HA:=evalm(2/(1+mu[pg])*[[1,-mu[pg]*z^(-1)]]*z^(Mp/2)&*QQp&*[[1],[Rp]])[1,1];
```

$$\mathrm{HA} := 2\,\frac{z^{(1/2\,\mathrm{Mp})}\,\left(\mathrm{P1}\,z^2 - z\,\mu_{\mathrm{pg}}\,\mathrm{P3} + \mathrm{Rp}\,\mathrm{P2}\,z - \mathrm{Rp}\,\mu_{\mathrm{pg}}\,\mathrm{P4}\right)}{(1+\mu_{\mathrm{pg}})\,z^2}$$

```
> Rp:=mu[pj]+(1+mu[pj])*(HB+HD);
```

$$\mathrm{Rp} := \mu_{\mathrm{pj}} + (1+\mu_{\mathrm{pj}})\,(\mathrm{HB} + \mathrm{HD})$$

```
> HE:=evalm([[0,1]]*z^(Mn/2-1)&*QQn&*[[1],[-mu[n1]]]*1/(1+mu[n1]))[1,1];
```

$$\mathrm{HE} := \frac{z^{(1/2\,\mathrm{Mn}-1)}\,(\mathrm{N3}\,z - \mathrm{N4}\,\mu_{\mathrm{n1}})}{(1+\mu_{\mathrm{n1}})\,z}$$

```
> HD:=(1+mu[nj])/(Rn-mu[nj])*(1+HB);
```

$$\mathrm{HD} := \frac{(1+\mu_{\mathrm{nj}})\,(1+\mathrm{HB})}{\dfrac{(\mathrm{N1}\,z - \mathrm{N2}\,\mu_{\mathrm{n1}})\,z}{\mathrm{N3}\,z - \mathrm{N4}\,\mu_{\mathrm{n1}}} - \mu_{\mathrm{nj}}}$$

```
> HC:=evalm([[0,1]]*z^(Mo/2-1)&*QQo&*[[1],[-mu[o1]]]*1/(1+mu[o1]))[1,1];
```

$$\mathrm{HC} := \frac{z^{(1/2\,\mathrm{Mo}-1)}\,(\mathrm{O3}\,z - \mathrm{O4}\,\mu_{\mathrm{o1}})}{(1+\mu_{\mathrm{o1}})\,z}$$

H_B is defined from equation (2.57):

```
> HB:=(1+mu[oj])*(Rn+1)/((Ro-mu[oj])*(Rn-mu[nj])-
(1+mu[oj])*(1+mu[nj]));
```

Finally H is defined from equation (2.64):

```
> H:=HB/HA/HC+HD/HA/HE;
```

This is the first representation of H. From this form the number of poles and zeros is not evident.

```
> HB:=(1+mu[oj])*(Rn+1)/((Ro-mu[oj])*(Rn-mu[nj])-(1+mu[oj])*(1+mu[nj]));
```

$$\mathrm{HB} := (1+\mu_{\mathrm{oj}})\left(\frac{(\mathrm{N1}\,z-\mathrm{N2}\,\mu_{\mathrm{n1}})\,z}{\mathrm{N3}\,z-\mathrm{N4}\,\mu_{\mathrm{n1}}}+1\right)\Big/\Bigg(\left(\frac{(\mathrm{O1}\,z-\mathrm{O2}\,\mu_{\mathrm{o1}})\,z}{\mathrm{O3}\,z-\mathrm{O4}\,\mu_{\mathrm{o1}}}-\mu_{\mathrm{oj}}\right)\left(\frac{(\mathrm{N1}\,z-\mathrm{N2}\,\mu_{\mathrm{n1}})\,z}{\mathrm{N3}\,z-\mathrm{N4}\,\mu_{\mathrm{n1}}}-\mu_{\mathrm{nj}}\right)-(1+\mu_{\mathrm{oj}})\,(1+\mu_{\mathrm{nj}})\Bigg)$$

```
> H:=HB/HA/HC+HD/HA/HE;
```

$$H := \frac{1}{2}(1+\mu_{\mathrm{oj}})\left(\frac{(\mathrm{N1}\,z-\mathrm{N2}\,\mu_{\mathrm{n1}})\,z}{\mathrm{N3}\,z-\mathrm{N4}\,\mu_{\mathrm{n1}}}+1\right)(1+\mu_{\mathrm{pg}})\,z^{3}\,(1+\mu_{\mathrm{o1}})\Big/\Bigg(\Bigg(\left(\frac{(\mathrm{O1}\,z-\mathrm{O2}\,\mu_{\mathrm{o1}})\,z}{\mathrm{O3}\,z-\mathrm{O4}\,\mu_{\mathrm{o1}}}-\mu_{\mathrm{oj}}\right)\left(\frac{(\mathrm{N1}\,z-\mathrm{N2}\,\mu_{\mathrm{n1}})\,z}{\mathrm{N3}\,z-\mathrm{N4}\,\mu_{\mathrm{n1}}}-\mu_{\mathrm{nj}}\right)-(1+\mu_{\mathrm{oj}})\,(1+\mu_{\mathrm{nj}})\Bigg)z^{(1/2\,\mathrm{Mp})}\Bigg(\mathrm{P1}\,z^{2}-z\,\mu_{\mathrm{pg}}\,\mathrm{P3}+\Bigg(\mu_{\mathrm{pj}}+(1+\mu_{\mathrm{pj}})\Bigg((1+\mu_{\mathrm{oj}})\left(\frac{(\mathrm{N1}\,z-\mathrm{N2}\,\mu_{\mathrm{n1}})\,z}{\mathrm{N3}\,z-\mathrm{N4}\,\mu_{\mathrm{n1}}}+1\right)\Big/\Bigg(\left(\frac{(\mathrm{O1}\,z-\mathrm{O2}\,\mu_{\mathrm{o1}})\,z}{\mathrm{O3}\,z-\mathrm{O4}\,\mu_{\mathrm{o1}}}-\mu_{\mathrm{oj}}\right)\left(\frac{(\mathrm{N1}\,z-\mathrm{N2}\,\mu_{\mathrm{n1}})\,z}{\mathrm{N3}\,z-\mathrm{N4}\,\mu_{\mathrm{n1}}}-\mu_{\mathrm{nj}}\right)-(1+\mu_{\mathrm{oj}})\,(1+\mu_{\mathrm{nj}})\Bigg)+(1+\mu_{\mathrm{nj}})\Bigg(1+(1+\mu_{\mathrm{oj}})\left(\frac{(\mathrm{N1}\,z-\mathrm{N2}\,\mu_{\mathrm{n1}})\,z}{\mathrm{N3}\,z-\mathrm{N4}\,\mu_{\mathrm{n1}}}+1\right)\Big/\Bigg(\left(\frac{(\mathrm{O1}\,z-\mathrm{O2}\,\mu_{\mathrm{o1}})\,z}{\mathrm{O3}\,z-\mathrm{O4}\,\mu_{\mathrm{o1}}}-\mu_{\mathrm{oj}}\right)\left(\frac{(\mathrm{N1}\,z-\mathrm{N2}\,\mu_{\mathrm{n1}})\,z}{\mathrm{N3}\,z-\mathrm{N4}\,\mu_{\mathrm{n1}}}-\mu_{\mathrm{nj}}\right)-(1+\mu_{\mathrm{oj}})\,(1+\mu_{\mathrm{nj}})\Bigg)\Bigg)\Big/\left(\frac{(\mathrm{N1}\,z-\mathrm{N2}\,\mu_{\mathrm{n1}})\,z}{\mathrm{N3}\,z-\mathrm{N4}\,\mu_{\mathrm{n1}}}-\mu_{\mathrm{nj}}\right)\Bigg)\Bigg)\mathrm{P2}\,z-\Bigg(\mu_{\mathrm{pj}}+(1+\mu_{\mathrm{pj}})\Bigg((1+\mu_{\mathrm{oj}})\left(\frac{(\mathrm{N1}\,z-\mathrm{N2}\,\mu_{\mathrm{n1}})\,z}{\mathrm{N3}\,z-\mathrm{N4}\,\mu_{\mathrm{n1}}}+1\right)\Big/\Bigg(\left(\frac{(\mathrm{O1}\,z-\mathrm{O2}\,\mu_{\mathrm{o1}})\,z}{\mathrm{O3}\,z-\mathrm{O4}\,\mu_{\mathrm{o1}}}-\mu_{\mathrm{oj}}\right)\left(\frac{(\mathrm{N1}\,z-\mathrm{N2}\,\mu_{\mathrm{n1}})\,z}{\mathrm{N3}\,z-\mathrm{N4}\,\mu_{\mathrm{n1}}}-\mu_{\mathrm{nj}}\right)-(1+\mu_{\mathrm{oj}})\,(1+\mu_{\mathrm{nj}})\Bigg)+(1+\mu_{\mathrm{nj}})\Bigg(1+(1+\mu_{\mathrm{oj}})\left(\frac{(\mathrm{N1}\,z-\mathrm{N2}\,\mu_{\mathrm{n1}})\,z}{\mathrm{N3}\,z-\mathrm{N4}\,\mu_{\mathrm{n1}}}+1\right)\Big/\Bigg(\left(\frac{(\mathrm{O1}\,z-\mathrm{O2}\,\mu_{\mathrm{o1}})\,z}{\mathrm{O3}\,z-\mathrm{O4}\,\mu_{\mathrm{o1}}}-\mu_{\mathrm{oj}}\right)\left(\frac{(\mathrm{N1}\,z-\mathrm{N2}\,\mu_{\mathrm{n1}})\,z}{\mathrm{N3}\,z-\mathrm{N4}\,\mu_{\mathrm{n1}}}-\mu_{\mathrm{nj}}\right)-(1+\mu_{\mathrm{oj}})\,(1+\mu_{\mathrm{nj}})\Bigg)\Bigg)\Big/\left(\frac{(\mathrm{N1}\,z-\mathrm{N2}\,\mu_{\mathrm{n1}})\,z}{\mathrm{N3}\,z-\mathrm{N4}\,\mu_{\mathrm{n1}}}-\mu_{\mathrm{nj}}\right)\Bigg)\Bigg)\mu_{\mathrm{pg}}\,\mathrm{P4}\Bigg)z^{(1/2\,\mathrm{Mo}-1)}\,(\mathrm{O3}\,z-\mathrm{O4}\,\mu_{\mathrm{o1}})\Bigg)+\frac{1}{2}(1+\mu_{\mathrm{nj}})\Bigg(1+(1+\mu_{\mathrm{oj}})\left(\frac{(\mathrm{N1}\,z-\mathrm{N2}\,\mu_{\mathrm{n1}})\,z}{\mathrm{N3}\,z-\mathrm{N4}\,\mu_{\mathrm{n1}}}+1\right)\Big/\Bigg(
$$

Here **MapleV** is requested to simplify H, which in this case essentially reorders the expression as a single (large) fraction.

Subsequently the `collect` function is called to isolate the coefficients of different powers of `z`.

```
> Hc:=collect(simplify(H), z);
```

$$\left(\frac{(\mathrm{O1}\,z-\mathrm{O2}\,\mu_{\mathrm{o1}})\;z}{\mathrm{O3}\,z-\mathrm{O4}\,\mu_{\mathrm{o1}}}-\mu_{\mathrm{oj}}\right)\;\left(\frac{(\mathrm{N1}\,z-\mathrm{N2}\,\mu_{\mathrm{n1}})\;z}{\mathrm{N3}\,z-\mathrm{N4}\,\mu_{\mathrm{n1}}}-\mu_{\mathrm{nj}}\right)$$
$$-\left(1+\mu_{\mathrm{oj}}\right)\;\left(1+\mu_{\mathrm{nj}}\right)\Bigg)\Bigg)\;\left(1+\mu_{\mathrm{pg}}\right)\;z^{3}\;\left(1+\mu_{\mathrm{n1}}\right)\Bigg/\Bigg($$
$$\left(\frac{(\mathrm{N1}\,z-\mathrm{N2}\,\mu_{\mathrm{n1}})\;z}{\mathrm{N3}\,z-\mathrm{N4}\,\mu_{\mathrm{n1}}}-\mu_{\mathrm{nj}}\right)\;z^{(\,1/2\,\mathrm{Mp}\,)}\left(\mathrm{P1}\,z^{2}-z\,\mu_{\mathrm{pg}}\,\mathrm{P3}+\left(\mu_{\mathrm{pj}}+\left(1+\mu_{\mathrm{pj}}\right)\left(\right.\right.\right.$$
$$\left(1+\mu_{\mathrm{oj}}\right)\;\left(\frac{(\mathrm{N1}\,z-\mathrm{N2}\,\mu_{\mathrm{n1}})\;z}{\mathrm{N3}\,z-\mathrm{N4}\,\mu_{\mathrm{n1}}}+1\right)\Bigg/\Bigg($$
$$\left(\frac{(\mathrm{O1}\,z-\mathrm{O2}\,\mu_{\mathrm{o1}})\;z}{\mathrm{O3}\,z-\mathrm{O4}\,\mu_{\mathrm{o1}}}-\mu_{\mathrm{oj}}\right)\;\left(\frac{(\mathrm{N1}\,z-\mathrm{N2}\,\mu_{\mathrm{n1}})\;z}{\mathrm{N3}\,z-\mathrm{N4}\,\mu_{\mathrm{n1}}}-\mu_{\mathrm{nj}}\right)$$
$$-\left(1+\mu_{\mathrm{oj}}\right)\;\left(1+\mu_{\mathrm{nj}}\right)\Bigg)\;+\left(1+\mu_{\mathrm{nj}}\right)\left(1+\left(1+\mu_{\mathrm{oj}}\right)\right.$$
$$\left(\frac{(\mathrm{N1}\,z-\mathrm{N2}\,\mu_{\mathrm{n1}})\;z}{\mathrm{N3}\,z-\mathrm{N4}\,\mu_{\mathrm{n1}}}+1\right)\Bigg/\Bigg($$
$$\left(\frac{(\mathrm{O1}\,z-\mathrm{O2}\,\mu_{\mathrm{o1}})\;z}{\mathrm{O3}\,z-\mathrm{O4}\,\mu_{\mathrm{o1}}}-\mu_{\mathrm{oj}}\right)\;\left(\frac{(\mathrm{N1}\,z-\mathrm{N2}\,\mu_{\mathrm{n1}})\;z}{\mathrm{N3}\,z-\mathrm{N4}\,\mu_{\mathrm{n1}}}-\mu_{\mathrm{nj}}\right)$$
$$-\left(1+\mu_{\mathrm{oj}}\right)\;\left(1+\mu_{\mathrm{nj}}\right)\Bigg)\Bigg)\Bigg/\left(\left(\frac{(\mathrm{N1}\,z-\mathrm{N2}\,\mu_{\mathrm{n1}})\;z}{\mathrm{N3}\,z-\mathrm{N4}\,\mu_{\mathrm{n1}}}-\mu_{\mathrm{nj}}\right)\right)\Bigg)\,\mathrm{P2}\,z-\left(\mu_{\mathrm{pj}}+\right.$$
$$\left(1+\mu_{\mathrm{pj}}\right)\left(\left(1+\mu_{\mathrm{oj}}\right)\;\left(\frac{(\mathrm{N1}\,z-\mathrm{N2}\,\mu_{\mathrm{n1}})\;z}{\mathrm{N3}\,z-\mathrm{N4}\,\mu_{\mathrm{n1}}}+1\right)\Bigg/\Bigg($$
$$\left(\frac{(\mathrm{O1}\,z-\mathrm{O2}\,\mu_{\mathrm{o1}})\;z}{\mathrm{O3}\,z-\mathrm{O4}\,\mu_{\mathrm{o1}}}-\mu_{\mathrm{oj}}\right)\;\left(\frac{(\mathrm{N1}\,z-\mathrm{N2}\,\mu_{\mathrm{n1}})\;z}{\mathrm{N3}\,z-\mathrm{N4}\,\mu_{\mathrm{n1}}}-\mu_{\mathrm{nj}}\right)$$
$$-\left(1+\mu_{\mathrm{oj}}\right)\;\left(1+\mu_{\mathrm{nj}}\right)\Bigg)\;+\left(1+\mu_{\mathrm{nj}}\right)\left(1+\left(1+\mu_{\mathrm{oj}}\right)\right.$$
$$\left(\frac{(\mathrm{N1}\,z-\mathrm{N2}\,\mu_{\mathrm{n1}})\;z}{\mathrm{N3}\,z-\mathrm{N4}\,\mu_{\mathrm{n1}}}+1\right)\Bigg/\Bigg($$
$$\left(\frac{(\mathrm{O1}\,z-\mathrm{O2}\,\mu_{\mathrm{o1}})\;z}{\mathrm{O3}\,z-\mathrm{O4}\,\mu_{\mathrm{o1}}}-\mu_{\mathrm{oj}}\right)\;\left(\frac{(\mathrm{N1}\,z-\mathrm{N2}\,\mu_{\mathrm{n1}})\;z}{\mathrm{N3}\,z-\mathrm{N4}\,\mu_{\mathrm{n1}}}-\mu_{\mathrm{nj}}\right)$$
$$-\left(1+\mu_{\mathrm{oj}}\right)\;\left(1+\mu_{\mathrm{nj}}\right)\Bigg)\Bigg)\Bigg/\left(\left(\frac{(\mathrm{N1}\,z-\mathrm{N2}\,\mu_{\mathrm{n1}})\;z}{\mathrm{N3}\,z-\mathrm{N4}\,\mu_{\mathrm{n1}}}-\mu_{\mathrm{nj}}\right)\right)\Bigg)\,\mu_{\mathrm{pg}}\,\mathrm{P4}\Bigg)$$
$$z^{(\,1/2\,\mathrm{Mn}-1\,)}\;\left(\mathrm{N3}\,z-\mathrm{N4}\,\mu_{\mathrm{n1}}\right)\Bigg)$$

```
> Hc:=collect(simplify(H), z);
```

$$\begin{aligned}\mathrm{Hc}:=\frac{1}{2}\Big(&-z^{(\,5-1/2\,\mathrm{Mp}-1/2\,\mathrm{Mo}\,)}\,\mu_{\mathrm{oj}}\,\mu_{\mathrm{o1}}\,\mathrm{N2}\,\mu_{\mathrm{n1}}-z^{(\,5-1/2\,\mathrm{Mp}-1/2\,\mathrm{Mo}\,)}\,\mu_{\mathrm{o1}}\,\mathrm{N2}\,\mu_{\mathrm{n1}}\\&-z^{(\,5-1/2\,\mathrm{Mp}-1/2\,\mathrm{Mn}\,)}\,\mu_{\mathrm{nj}}\,\mathrm{O2}\,\mu_{\mathrm{o1}}-z^{(\,5-1/2\,\mathrm{Mp}-1/2\,\mathrm{Mn}\,)}\,\mathrm{O2}\,\mu_{\mathrm{o1}}\,\mu_{\mathrm{n1}}\\&+z^{(\,5-1/2\,\mathrm{Mp}-1/2\,\mathrm{Mo}\,)}\,\mu_{\mathrm{oj}}\,\mathrm{N3}-z^{(\,-1/2\,\mathrm{Mp}+4-1/2\,\mathrm{Mo}\,)}\,\mathrm{N4}\,\mu_{\mathrm{n1}}\\&-z^{(\,-1/2\,\mathrm{Mp}+4-1/2\,\mathrm{Mn}\,)}\,\mathrm{O4}\,\mu_{\mathrm{o1}}-z^{(\,5-1/2\,\mathrm{Mp}-1/2\,\mathrm{Mo}\,)}\,\mathrm{N2}\,\mu_{\mathrm{n1}}\\&+z^{(\,5-1/2\,\mathrm{Mp}-1/2\,\mathrm{Mn}\,)}\,\mu_{\mathrm{nj}}\,\mathrm{O3}+z^{(\,5-1/2\,\mathrm{Mp}-1/2\,\mathrm{Mn}\,)}\,\mathrm{O3}\,\mu_{\mathrm{n1}}\\&-z^{(\,5-1/2\,\mathrm{Mp}-1/2\,\mathrm{Mn}\,)}\,\mu_{\mathrm{nj}}\,\mathrm{O2}\,\mu_{\mathrm{o1}}\,\mu_{\mathrm{n1}}+\mathrm{N1}\,z^{(\,6-1/2\,\mathrm{Mp}-1/2\,\mathrm{Mo}\,)}\,\mu_{\mathrm{o1}}\\&+\mathrm{N1}\,z^{(\,6-1/2\,\mathrm{Mp}-1/2\,\mathrm{Mo}\,)}\,\mu_{\mathrm{oj}}\,\mu_{\mathrm{o1}}+\mathrm{N1}\,z^{(\,6-1/2\,\mathrm{Mp}-1/2\,\mathrm{Mo}\,)}\,\mu_{\mathrm{oj}}\\&+z^{(\,6-1/2\,\mathrm{Mp}-1/2\,\mathrm{Mn}\,)}\,\mu_{\mathrm{nj}}\,\mathrm{O1}\,\mu_{\mathrm{n1}}+z^{(\,6-1/2\,\mathrm{Mp}-1/2\,\mathrm{Mn}\,)}\,\mu_{\mathrm{nj}}\,\mathrm{O1}\\&+z^{(\,5-1/2\,\mathrm{Mp}-1/2\,\mathrm{Mo}\,)}\,\mathrm{N3}\,\mu_{\mathrm{o1}}-z^{(\,-1/2\,\mathrm{Mp}+4-1/2\,\mathrm{Mo}\,)}\,\mu_{\mathrm{oj}}\,\mathrm{N4}\,\mu_{\mathrm{n1}}\end{aligned}$$

Notice the fraction line separating the numerator and denominator of the transfer function.

$$
\begin{aligned}
&- z^{(-1/2\,\mathrm{Mp}+4-1/2\,\mathrm{Mn})}\,\mu_{\mathrm{nj}}\,\mathrm{O4}\,\mu_{\mathrm{o1}}\,\mu_{\mathrm{n1}} - z^{(-1/2\,\mathrm{Mp}+4-1/2\,\mathrm{Mo})}\,\mu_{\mathrm{oj}}\,\mathrm{N4}\,\mu_{\mathrm{n1}}\,\mu_{\mathrm{o1}}\\
&- z^{(-1/2\,\mathrm{Mp}+4-1/2\,\mathrm{Mo})}\,\mathrm{N4}\,\mu_{\mathrm{n1}}\,\mu_{\mathrm{o1}} - z^{(-1/2\,\mathrm{Mp}+4-1/2\,\mathrm{Mn})}\,\mathrm{O4}\,\mu_{\mathrm{o1}}\,\mu_{\mathrm{n1}}\\
&- z^{(-1/2\,\mathrm{Mp}+4-1/2\,\mathrm{Mn})}\,\mu_{\mathrm{nj}}\,\mathrm{O4}\,\mu_{\mathrm{o1}} - z^{(5-1/2\,\mathrm{Mp}-1/2\,\mathrm{Mo})}\,\mu_{\mathrm{oj}}\,\mathrm{N2}\,\mu_{\mathrm{n1}}\\
&+ \mathrm{N1}\,z^{(6-1/2\,\mathrm{Mp}-1/2\,\mathrm{Mo})} + z^{(5-1/2\,\mathrm{Mp}-1/2\,\mathrm{Mn})}\,\mathrm{O3}\\
&+ z^{(5-1/2\,\mathrm{Mp}-1/2\,\mathrm{Mn})}\,\mu_{\mathrm{nj}}\,\mathrm{O3}\,\mu_{\mathrm{n1}} - z^{(5-1/2\,\mathrm{Mp}-1/2\,\mathrm{Mn})}\,\mathrm{O2}\,\mu_{\mathrm{o1}}\\
&+ z^{(6-1/2\,\mathrm{Mp}-1/2\,\mathrm{Mn})}\,\mathrm{O1} + z^{(5-1/2\,\mathrm{Mp}-1/2\,\mathrm{Mo})}\,\mathrm{N3}\\
&+ z^{(5-1/2\,\mathrm{Mp}-1/2\,\mathrm{Mo})}\,\mu_{\mathrm{oj}}\,\mathrm{N3}\,\mu_{\mathrm{o1}} + z^{(6-1/2\,\mathrm{Mp}-1/2\,\mathrm{Mn})}\,\mathrm{O1}\,\mu_{\mathrm{n1}}\Big)\,(1+\mu_{\mathrm{pg}})\Big/\Big(\\
&\mathrm{N1}\,\mathrm{P1}\,\mathrm{O1}\,z^6 + \Big(-\mathrm{N1}\,\mu_{\mathrm{pg}}\,\mathrm{P3}\,\mathrm{O1} - \mathrm{N1}\,\mathrm{P1}\,\mathrm{O2}\,\mu_{\mathrm{o1}} - \mathrm{P1}\,\mathrm{O1}\,\mathrm{N2}\,\mu_{\mathrm{n1}}\\
&- \mathrm{P1}\,\mu_{\mathrm{nj}}\,\mathrm{N3}\,\mathrm{O1} + \mathrm{N1}\,\mathrm{P2}\,\mu_{\mathrm{pj}}\,\mathrm{O1} - \mathrm{N1}\,\mathrm{P1}\,\mu_{\mathrm{oj}}\,\mathrm{O3}\Big)\,z^5 + \Big(-\mathrm{P1}\,\mathrm{O3}\,\mu_{\mathrm{nj}}\,\mathrm{N3}\\
&+ \mathrm{P1}\,\mu_{\mathrm{nj}}\,\mathrm{N4}\,\mu_{\mathrm{n1}}\,\mathrm{O1} - \mathrm{P1}\,\mu_{\mathrm{oj}}\,\mathrm{O3}\,\mathrm{N3} + \mathrm{N1}\,\mathrm{P1}\,\mu_{\mathrm{oj}}\,\mathrm{O4}\,\mu_{\mathrm{o1}} + \mathrm{P2}\,\mathrm{N3}\,\mathrm{O1}\\
&- \mathrm{P1}\,\mathrm{O3}\,\mathrm{N3} + \mathrm{P1}\,\mathrm{O2}\,\mu_{\mathrm{o1}}\,\mathrm{N2}\,\mu_{\mathrm{n1}} - \mathrm{P2}\,\mu_{\mathrm{pj}}\,\mathrm{O1}\,\mathrm{N2}\,\mu_{\mathrm{n1}} + \mathrm{P2}\,\mu_{\mathrm{pj}}\,\mathrm{N3}\,\mathrm{O1}\\
&+ \mathrm{P2}\,\mu_{\mathrm{nj}}\,\mathrm{N3}\,\mathrm{O1} + \mu_{\mathrm{pg}}\,\mathrm{P3}\,\mathrm{O1}\,\mathrm{N2}\,\mu_{\mathrm{n1}} + \mathrm{P1}\,\mu_{\mathrm{oj}}\,\mathrm{O3}\,\mathrm{N2}\,\mu_{\mathrm{n1}}\\
&+ \mu_{\mathrm{pg}}\,\mathrm{P3}\,\mu_{\mathrm{nj}}\,\mathrm{N3}\,\mathrm{O1} + \mathrm{N1}\,\mathrm{P2}\,\mu_{\mathrm{oj}}\,\mathrm{O3} + \mathrm{N1}\,\mathrm{P2}\,\mathrm{O3} + \mathrm{N1}\,\mu_{\mathrm{pg}}\,\mathrm{P3}\,\mathrm{O2}\,\mu_{\mathrm{o1}}\\
&- \mathrm{N1}\,\mu_{\mathrm{pg}}\,\mathrm{P4}\,\mu_{\mathrm{pj}}\,\mathrm{O1} + \mathrm{N1}\,\mu_{\mathrm{pg}}\,\mathrm{P3}\,\mu_{\mathrm{oj}}\,\mathrm{O3} + \mathrm{P1}\,\mu_{\mathrm{nj}}\,\mathrm{N3}\,\mathrm{O2}\,\mu_{\mathrm{o1}}\\
&+ \mathrm{N1}\,\mathrm{P2}\,\mu_{\mathrm{pj}}\,\mathrm{O3} - \mathrm{N1}\,\mathrm{P2}\,\mu_{\mathrm{pj}}\,\mathrm{O2}\,\mu_{\mathrm{o1}}\Big)\,z^4 + \Big(-\mathrm{N1}\,\mathrm{P2}\,\mu_{\mathrm{pj}}\,\mathrm{O4}\,\mu_{\mathrm{o1}}\\
&- \mathrm{N1}\,\mu_{\mathrm{pg}}\,\mathrm{P4}\,\mu_{\mathrm{oj}}\,\mathrm{O3} - \mathrm{N1}\,\mu_{\mathrm{pg}}\,\mathrm{P4}\,\mu_{\mathrm{pj}}\,\mathrm{O3} - \mathrm{P2}\,\mathrm{N4}\,\mu_{\mathrm{n1}}\,\mathrm{O1} + \mathrm{P2}\,\mu_{\mathrm{oj}}\,\mathrm{O3}\,\mathrm{N3}\\
&- \mathrm{P2}\,\mu_{\mathrm{nj}}\,\mathrm{N3}\,\mathrm{O2}\,\mu_{\mathrm{o1}} + 2\,\mathrm{P2}\,\mathrm{O3}\,\mathrm{N3} - \mathrm{P2}\,\mu_{\mathrm{pj}}\,\mathrm{O3}\,\mathrm{N2}\,\mu_{\mathrm{n1}} - \mu_{\mathrm{pg}}\,\mathrm{P4}\,\mathrm{N3}\,\mathrm{O1}\\
&- \mathrm{P2}\,\mu_{\mathrm{nj}}\,\mathrm{N4}\,\mu_{\mathrm{n1}}\,\mathrm{O1} + \mu_{\mathrm{pg}}\,\mathrm{P3}\,\mathrm{O3}\,\mu_{\mathrm{nj}}\,\mathrm{N3} + \mu_{\mathrm{pg}}\,\mathrm{P3}\,\mathrm{O3}\,\mathrm{N3}\\
&+ \mathrm{P1}\,\mu_{\mathrm{oj}}\,\mathrm{O3}\,\mathrm{N4}\,\mu_{\mathrm{n1}} + \mathrm{P1}\,\mu_{\mathrm{oj}}\,\mathrm{O4}\,\mu_{\mathrm{o1}}\,\mathrm{N3} - \mathrm{P2}\,\mu_{\mathrm{pj}}\,\mathrm{N4}\,\mu_{\mathrm{n1}}\,\mathrm{O1}\\
&- \mu_{\mathrm{pg}}\,\mathrm{P4}\,\mu_{\mathrm{pj}}\,\mathrm{N3}\,\mathrm{O1} - \mu_{\mathrm{pg}}\,\mathrm{P4}\,\mu_{\mathrm{nj}}\,\mathrm{N3}\,\mathrm{O1} + \mathrm{P2}\,\mu_{\mathrm{pj}}\,\mathrm{O2}\,\mu_{\mathrm{o1}}\,\mathrm{N2}\,\mu_{\mathrm{n1}}\\
&- \mathrm{P1}\,\mu_{\mathrm{oj}}\,\mathrm{O4}\,\mu_{\mathrm{o1}}\,\mathrm{N2}\,\mu_{\mathrm{n1}} + \mathrm{P1}\,\mathrm{O4}\,\mu_{\mathrm{o1}}\,\mu_{\mathrm{nj}}\,\mathrm{N3} + \mathrm{P1}\,\mathrm{O3}\,\mathrm{N4}\,\mu_{\mathrm{n1}}\\
&- \mu_{\mathrm{pg}}\,\mathrm{P3}\,\mu_{\mathrm{oj}}\,\mathrm{O3}\,\mathrm{N2}\,\mu_{\mathrm{n1}} - \mathrm{P2}\,\mathrm{O3}\,\mathrm{N2}\,\mu_{\mathrm{n1}} - \mathrm{P1}\,\mu_{\mathrm{nj}}\,\mathrm{N4}\,\mu_{\mathrm{n1}}\,\mathrm{O2}\,\mu_{\mathrm{o1}}\\
&- \mathrm{P2}\,\mathrm{N3}\,\mathrm{O2}\,\mu_{\mathrm{o1}} - \mathrm{N1}\,\mu_{\mathrm{pg}}\,\mathrm{P4}\,\mathrm{O3} - \mathrm{N1}\,\mu_{\mathrm{pg}}\,\mathrm{P3}\,\mu_{\mathrm{oj}}\,\mathrm{O4}\,\mu_{\mathrm{o1}}\\
&+ \mathrm{N1}\,\mu_{\mathrm{pg}}\,\mathrm{P4}\,\mu_{\mathrm{pj}}\,\mathrm{O2}\,\mu_{\mathrm{o1}} - \mathrm{N1}\,\mathrm{P2}\,\mathrm{O4}\,\mu_{\mathrm{o1}} - \mathrm{P2}\,\mu_{\mathrm{oj}}\,\mathrm{O3}\,\mathrm{N2}\,\mu_{\mathrm{n1}}\\
&- \mu_{\mathrm{pg}}\,\mathrm{P3}\,\mu_{\mathrm{nj}}\,\mathrm{N3}\,\mathrm{O2}\,\mu_{\mathrm{o1}} - \mu_{\mathrm{pg}}\,\mathrm{P3}\,\mu_{\mathrm{nj}}\,\mathrm{N4}\,\mu_{\mathrm{n1}}\,\mathrm{O1} - \mu_{\mathrm{pg}}\,\mathrm{P3}\,\mathrm{O2}\,\mu_{\mathrm{o1}}\,\mathrm{N2}\,\mu_{\mathrm{n1}}\\
&+ \mathrm{P2}\,\mu_{\mathrm{pj}}\,\mathrm{O3}\,\mathrm{N3} + \mathrm{P2}\,\mathrm{O3}\,\mu_{\mathrm{nj}}\,\mathrm{N3} + \mathrm{P1}\,\mathrm{O4}\,\mu_{\mathrm{o1}}\,\mathrm{N3}\\
&+ \mu_{\mathrm{pg}}\,\mathrm{P4}\,\mu_{\mathrm{pj}}\,\mathrm{O1}\,\mathrm{N2}\,\mu_{\mathrm{n1}} + \mathrm{P1}\,\mathrm{O3}\,\mu_{\mathrm{nj}}\,\mathrm{N4}\,\mu_{\mathrm{n1}} + \mu_{\mathrm{pg}}\,\mathrm{P3}\,\mu_{\mathrm{oj}}\,\mathrm{O3}\,\mathrm{N3}\\
&- \mathrm{P2}\,\mu_{\mathrm{pj}}\,\mathrm{N3}\,\mathrm{O2}\,\mu_{\mathrm{o1}} - \mathrm{N1}\,\mathrm{P2}\,\mu_{\mathrm{oj}}\,\mathrm{O4}\,\mu_{\mathrm{o1}}\Big)\,z^3 + \Big(-\mu_{\mathrm{pg}}\,\mathrm{P4}\,\mu_{\mathrm{pj}}\,\mathrm{O2}\,\mu_{\mathrm{o1}}\,\mathrm{N2}\,\mu_{\mathrm{n1}}\\
&+ \mu_{\mathrm{pg}}\,\mathrm{P4}\,\mu_{\mathrm{pj}}\,\mathrm{O3}\,\mathrm{N2}\,\mu_{\mathrm{n1}} + \mu_{\mathrm{pg}}\,\mathrm{P4}\,\mu_{\mathrm{pj}}\,\mathrm{N3}\,\mathrm{O2}\,\mu_{\mathrm{o1}} + \mathrm{P2}\,\mu_{\mathrm{pj}}\,\mathrm{N4}\,\mu_{\mathrm{n1}}\,\mathrm{O2}\,\mu_{\mathrm{o1}}\\
&+ \mathrm{N1}\,\mu_{\mathrm{pg}}\,\mathrm{P4}\,\mathrm{O4}\,\mu_{\mathrm{o1}} + \mathrm{N1}\,\mu_{\mathrm{pg}}\,\mathrm{P4}\,\mu_{\mathrm{oj}}\,\mathrm{O4}\,\mu_{\mathrm{o1}} + \mathrm{N1}\,\mu_{\mathrm{pg}}\,\mathrm{P4}\,\mu_{\mathrm{pj}}\,\mathrm{O4}\,\mu_{\mathrm{o1}}\\
&- \mu_{\mathrm{pg}}\,\mathrm{P3}\,\mathrm{O3}\,\mu_{\mathrm{nj}}\,\mathrm{N4}\,\mu_{\mathrm{n1}} + \mu_{\mathrm{pg}}\,\mathrm{P4}\,\mu_{\mathrm{oj}}\,\mathrm{O3}\,\mathrm{N2}\,\mu_{\mathrm{n1}} - \mu_{\mathrm{pg}}\,\mathrm{P4}\,\mu_{\mathrm{oj}}\,\mathrm{O3}\,\mathrm{N3}\\
&- \mathrm{P2}\,\mu_{\mathrm{pj}}\,\mathrm{O3}\,\mathrm{N4}\,\mu_{\mathrm{n1}} + \mu_{\mathrm{pg}}\,\mathrm{P4}\,\mu_{\mathrm{pj}}\,\mathrm{N4}\,\mu_{\mathrm{n1}}\,\mathrm{O1} + \mu_{\mathrm{pg}}\,\mathrm{P4}\,\mu_{\mathrm{nj}}\,\mathrm{N3}\,\mathrm{O2}\,\mu_{\mathrm{o1}}\\
&+ \mu_{\mathrm{pg}}\,\mathrm{P4}\,\mu_{\mathrm{nj}}\,\mathrm{N4}\,\mu_{\mathrm{n1}}\,\mathrm{O1} + \mathrm{P2}\,\mu_{\mathrm{pj}}\,\mathrm{O4}\,\mu_{\mathrm{o1}}\,\mathrm{N2}\,\mu_{\mathrm{n1}} + \mu_{\mathrm{pg}}\,\mathrm{P4}\,\mathrm{O3}\,\mathrm{N2}\,\mu_{\mathrm{n1}}\\
&+ \mu_{\mathrm{pg}}\,\mathrm{P4}\,\mathrm{N3}\,\mathrm{O2}\,\mu_{\mathrm{o1}} + \mu_{\mathrm{pg}}\,\mathrm{P4}\,\mathrm{N4}\,\mu_{\mathrm{n1}}\,\mathrm{O1} + \mathrm{P2}\,\mu_{\mathrm{nj}}\,\mathrm{N4}\,\mu_{\mathrm{n1}}\,\mathrm{O2}\,\mu_{\mathrm{o1}}\\
&- \mu_{\mathrm{pg}}\,\mathrm{P3}\,\mu_{\mathrm{oj}}\,\mathrm{O3}\,\mathrm{N4}\,\mu_{\mathrm{n1}} - 2\,\mathrm{P2}\,\mathrm{O4}\,\mu_{\mathrm{o1}}\,\mathrm{N3} - 2\,\mathrm{P2}\,\mathrm{O3}\,\mathrm{N4}\,\mu_{\mathrm{n1}}\\
&+ \mu_{\mathrm{pg}}\,\mathrm{P3}\,\mu_{\mathrm{nj}}\,\mathrm{N4}\,\mu_{\mathrm{n1}}\,\mathrm{O2}\,\mu_{\mathrm{o1}} + \mathrm{P2}\,\mathrm{N4}\,\mu_{\mathrm{n1}}\,\mathrm{O2}\,\mu_{\mathrm{o1}} - \mu_{\mathrm{pg}}\,\mathrm{P3}\,\mu_{\mathrm{oj}}\,\mathrm{O4}\,\mu_{\mathrm{o1}}\,\mathrm{N3}\\
&- \mathrm{P2}\,\mu_{\mathrm{oj}}\,\mathrm{O3}\,\mathrm{N4}\,\mu_{\mathrm{n1}} - \mu_{\mathrm{pg}}\,\mathrm{P3}\,\mathrm{O4}\,\mu_{\mathrm{o1}}\,\mu_{\mathrm{nj}}\,\mathrm{N3} - \mathrm{P1}\,\mu_{\mathrm{oj}}\,\mathrm{O4}\,\mu_{\mathrm{o1}}\,\mathrm{N4}\,\mu_{\mathrm{n1}}\\
&+ \mathrm{P2}\,\mu_{\mathrm{oj}}\,\mathrm{O4}\,\mu_{\mathrm{o1}}\,\mathrm{N2}\,\mu_{\mathrm{n1}} - \mu_{\mathrm{pg}}\,\mathrm{P3}\,\mathrm{O4}\,\mu_{\mathrm{o1}}\,\mathrm{N3} - \mu_{\mathrm{pg}}\,\mathrm{P3}\,\mathrm{O3}\,\mathrm{N4}\,\mu_{\mathrm{n1}}\\
&+ \mu_{\mathrm{pg}}\,\mathrm{P3}\,\mu_{\mathrm{oj}}\,\mathrm{O4}\,\mu_{\mathrm{o1}}\,\mathrm{N2}\,\mu_{\mathrm{n1}} + \mathrm{P2}\,\mathrm{O4}\,\mu_{\mathrm{o1}}\,\mathrm{N2}\,\mu_{\mathrm{n1}} - 2\,\mu_{\mathrm{pg}}\,\mathrm{P4}\,\mathrm{O3}\,\mathrm{N3}\\
&- \mathrm{P2}\,\mu_{\mathrm{oj}}\,\mathrm{O4}\,\mu_{\mathrm{o1}}\,\mathrm{N3} - \mathrm{P2}\,\mu_{\mathrm{pj}}\,\mathrm{O4}\,\mu_{\mathrm{o1}}\,\mathrm{N3} - \mathrm{P2}\,\mathrm{O3}\,\mu_{\mathrm{nj}}\,\mathrm{N4}\,\mu_{\mathrm{n1}}\\
&- \mathrm{P1}\,\mathrm{O4}\,\mu_{\mathrm{o1}}\,\mathrm{N4}\,\mu_{\mathrm{n1}} - \mu_{\mathrm{pg}}\,\mathrm{P4}\,\mathrm{O3}\,\mu_{\mathrm{nj}}\,\mathrm{N3} - \mathrm{P2}\,\mathrm{O4}\,\mu_{\mathrm{o1}}\,\mu_{\mathrm{nj}}\,\mathrm{N3}\\
&- \mathrm{P1}\,\mathrm{O4}\,\mu_{\mathrm{o1}}\,\mu_{\mathrm{nj}}\,\mathrm{N4}\,\mu_{\mathrm{n1}} - \mu_{\mathrm{pg}}\,\mathrm{P4}\,\mu_{\mathrm{pj}}\,\mathrm{O3}\,\mathrm{N3}\Big)\,z^2 + \Big(
\end{aligned}
$$

The denominator of the previous expression is the first suboperand of the fourth suboperand of `Hc` and is denoted `dHc`:

```
> dHc:= op(1,op(4,Hc)):  # Denominator of Hc
```

The following lines extract the coefficients f_i of all the powers of z^i in the denominator:

```
> for i from 0 to degree(dHc,z) do
>  f[i]:=coeff(dHc,z,i);
> od;
```

$$
\begin{aligned}
&-\mu_{pg}\,P4\,\mu_{oj}\,O4\,\mu_{o1}\,N2\,\mu_{n1}+P2\,\mu_{pj}\,O4\,\mu_{o1}\,N4\,\mu_{n1}+2\,\mu_{pg}\,P4\,O3\,N4\,\mu_{n1}\\
&+2\,\mu_{pg}\,P4\,O4\,\mu_{o1}\,N3+\mu_{pg}\,P3\,O4\,\mu_{o1}\,N4\,\mu_{n1}+\mu_{pg}\,P4\,\mu_{pj}\,O3\,N4\,\mu_{n1}\\
&+\mu_{pg}\,P4\,O3\,\mu_{nj}\,N4\,\mu_{n1}-\mu_{pg}\,P4\,O4\,\mu_{o1}\,N2\,\mu_{n1}+P2\,O4\,\mu_{o1}\,\mu_{nj}\,N4\,\mu_{n1}\\
&+\mu_{pg}\,P4\,O4\,\mu_{o1}\,\mu_{nj}\,N3-\mu_{pg}\,P4\,\mu_{pj}\,N4\,\mu_{n1}\,O2\,\mu_{o1}\\
&+\mu_{pg}\,P4\,\mu_{oj}\,O3\,N4\,\mu_{n1}-\mu_{pg}\,P4\,N4\,\mu_{n1}\,O2\,\mu_{o1}\\
&-\mu_{pg}\,P4\,\mu_{nj}\,N4\,\mu_{n1}\,O2\,\mu_{o1}+2\,P2\,O4\,\mu_{o1}\,N4\,\mu_{n1}\\
&-\mu_{pg}\,P4\,\mu_{pj}\,O4\,\mu_{o1}\,N2\,\mu_{n1}+P2\,\mu_{oj}\,O4\,\mu_{o1}\,N4\,\mu_{n1}\\
&+\mu_{pg}\,P4\,\mu_{pj}\,O4\,\mu_{o1}\,N3+\mu_{pg}\,P3\,\mu_{oj}\,O4\,\mu_{o1}\,N4\,\mu_{n1}\\
&+\mu_{pg}\,P4\,\mu_{oj}\,O4\,\mu_{o1}\,N3+\mu_{pg}\,P3\,O4\,\mu_{o1}\,\mu_{nj}\,N4\,\mu_{n1}\Big)\,z\\
&-\mu_{pg}\,P4\,\mu_{pj}\,O4\,\mu_{o1}\,N4\,\mu_{n1}-\mu_{pg}\,P4\,O4\,\mu_{o1}\,\mu_{nj}\,N4\,\mu_{n1}\\
&-2\,\mu_{pg}\,P4\,O4\,\mu_{o1}\,N4\,\mu_{n1}-\mu_{pg}\,P4\,\mu_{oj}\,O4\,\mu_{o1}\,N4\,\mu_{n1}\Big)
\end{aligned}
$$

```
> dHc:= op(1,op(4,Hc)):  # Denominator of Hc
> for i from 0 to degree(dHc,z) do
>  f[i]:=coeff(dHc,z,i);
> od;
```

$$
\begin{aligned}
f_0 :=\ &-\mu_{pg}\,P4\,\mu_{pj}\,O4\,\mu_{o1}\,N4\,\mu_{n1}-\mu_{pg}\,P4\,O4\,\mu_{o1}\,\mu_{nj}\,N4\,\mu_{n1}\\
&-2\,\mu_{pg}\,P4\,O4\,\mu_{o1}\,N4\,\mu_{n1}-\mu_{pg}\,P4\,\mu_{oj}\,O4\,\mu_{o1}\,N4\,\mu_{n1}
\end{aligned}
$$

$$
\begin{aligned}
f_1 :=\ &-\mu_{pg}\,P4\,\mu_{oj}\,O4\,\mu_{o1}\,N2\,\mu_{n1}+P2\,\mu_{pj}\,O4\,\mu_{o1}\,N4\,\mu_{n1}+2\,\mu_{pg}\,P4\,O3\,N4\,\mu_{n1}\\
&+2\,\mu_{pg}\,P4\,O4\,\mu_{o1}\,N3+\mu_{pg}\,P3\,O4\,\mu_{o1}\,N4\,\mu_{n1}+\mu_{pg}\,P4\,\mu_{pj}\,O3\,N4\,\mu_{n1}\\
&+\mu_{pg}\,P4\,O3\,\mu_{nj}\,N4\,\mu_{n1}-\mu_{pg}\,P4\,O4\,\mu_{o1}\,N2\,\mu_{n1}+P2\,O4\,\mu_{o1}\,\mu_{nj}\,N4\,\mu_{n1}\\
&+\mu_{pg}\,P4\,O4\,\mu_{o1}\,\mu_{nj}\,N3-\mu_{pg}\,P4\,\mu_{pj}\,N4\,\mu_{n1}\,O2\,\mu_{o1}\\
&+\mu_{pg}\,P4\,\mu_{oj}\,O3\,N4\,\mu_{n1}-\mu_{pg}\,P4\,N4\,\mu_{n1}\,O2\,\mu_{o1}\\
&-\mu_{pg}\,P4\,\mu_{nj}\,N4\,\mu_{n1}\,O2\,\mu_{o1}+2\,P2\,O4\,\mu_{o1}\,N4\,\mu_{n1}\\
&-\mu_{pg}\,P4\,\mu_{pj}\,O4\,\mu_{o1}\,N2\,\mu_{n1}+P2\,\mu_{oj}\,O4\,\mu_{o1}\,N4\,\mu_{n1}\\
&+\mu_{pg}\,P4\,\mu_{pj}\,O4\,\mu_{o1}\,N3+\mu_{pg}\,P3\,\mu_{oj}\,O4\,\mu_{o1}\,N4\,\mu_{n1}\\
&+\mu_{pg}\,P4\,\mu_{oj}\,O4\,\mu_{o1}\,N3+\mu_{pg}\,P3\,O4\,\mu_{o1}\,\mu_{nj}\,N4\,\mu_{n1}
\end{aligned}
$$

$$
\begin{aligned}
f_2 :=\ &-\mu_{pg}\,P4\,\mu_{pj}\,O2\,\mu_{o1}\,N2\,\mu_{n1}+\mu_{pg}\,P4\,\mu_{pj}\,O3\,N2\,\mu_{n1}+\mu_{pg}\,P4\,\mu_{pj}\,N3\,O2\,\mu_{o1}\\
&+P2\,\mu_{pj}\,N4\,\mu_{n1}\,O2\,\mu_{o1}+N1\,\mu_{pg}\,P4\,O4\,\mu_{o1}+N1\,\mu_{pg}\,P4\,\mu_{oj}\,O4\,\mu_{o1}\\
&+N1\,\mu_{pg}\,P4\,\mu_{pj}\,O4\,\mu_{o1}-\mu_{pg}\,P3\,O3\,\mu_{nj}\,N4\,\mu_{n1}+\mu_{pg}\,P4\,\mu_{oj}\,O3\,N2\,\mu_{n1}\\
&-\mu_{pg}\,P4\,\mu_{oj}\,O3\,N3-P2\,\mu_{pj}\,O3\,N4\,\mu_{n1}+\mu_{pg}\,P4\,\mu_{pj}\,N4\,\mu_{n1}\,O1\\
&+\mu_{pg}\,P4\,\mu_{nj}\,N3\,O2\,\mu_{o1}+\mu_{pg}\,P4\,\mu_{nj}\,N4\,\mu_{n1}\,O1+P2\,\mu_{pj}\,O4\,\mu_{o1}\,N2\,\mu_{n1}\\
&+\mu_{pg}\,P4\,O3\,N2\,\mu_{n1}+\mu_{pg}\,P4\,N3\,O2\,\mu_{o1}+\mu_{pg}\,P4\,N4\,\mu_{n1}\,O1\\
&+P2\,\mu_{nj}\,N4\,\mu_{n1}\,O2\,\mu_{o1}-\mu_{pg}\,P3\,\mu_{oj}\,O3\,N4\,\mu_{n1}-2\,P2\,O4\,\mu_{o1}\,N3\\
&-2\,P2\,O3\,N4\,\mu_{n1}+\mu_{pg}\,P3\,\mu_{nj}\,N4\,\mu_{n1}\,O2\,\mu_{o1}+P2\,N4\,\mu_{n1}\,O2\,\mu_{o1}\\
&-\mu_{pg}\,P3\,\mu_{oj}\,O4\,\mu_{o1}\,N3-P2\,\mu_{oj}\,O3\,N4\,\mu_{n1}-\mu_{pg}\,P3\,O4\,\mu_{o1}\,\mu_{nj}\,N3\\
&-P1\,\mu_{oj}\,O4\,\mu_{o1}\,N4\,\mu_{n1}+P2\,\mu_{oj}\,O4\,\mu_{o1}\,N2\,\mu_{n1}-\mu_{pg}\,P3\,O4\,\mu_{o1}\,N3\\
&-\mu_{pg}\,P3\,O3\,N4\,\mu_{n1}+\mu_{pg}\,P3\,\mu_{oj}\,O4\,\mu_{o1}\,N2\,\mu_{n1}+P2\,O4\,\mu_{o1}\,N2\,\mu_{n1}\\
&-2\,\mu_{pg}\,P4\,O3\,N3-P2\,\mu_{oj}\,O4\,\mu_{o1}\,N3-P2\,\mu_{pj}\,O4\,\mu_{o1}\,N3\\
&-P2\,O3\,\mu_{nj}\,N4\,\mu_{n1}-P1\,O4\,\mu_{o1}\,N4\,\mu_{n1}-\mu_{pg}\,P4\,O3\,\mu_{nj}\,N3\\
&-P2\,O4\,\mu_{o1}\,\mu_{nj}\,N3-P1\,O4\,\mu_{o1}\,\mu_{nj}\,N4\,\mu_{n1}-\mu_{pg}\,P4\,\mu_{pj}\,O3\,N3
\end{aligned}
$$

$$
\begin{aligned}
f_3 :=\ &-N1\,P2\,\mu_{pj}\,O4\,\mu_{o1}-N1\,\mu_{pg}\,P4\,\mu_{oj}\,O3-N1\,\mu_{pg}\,P4\,\mu_{pj}\,O3-P2\,N4\,\mu_{n1}\,O1\\
&+P2\,\mu_{oj}\,O3\,N3-P2\,\mu_{nj}\,N3\,O2\,\mu_{o1}+2\,P2\,O3\,N3-P2\,\mu_{pj}\,O3\,N2\,\mu_{n1}
\end{aligned}
$$

The denominator of the transfer function can be expressed as

$$f_6 z^6 + f_5 z^5 + f_4 z^4 + f_3 z^3 + f_2 z^2 + f_1 z + f_0 \tag{A.1}$$

where all f_i's are linear combinations of products of three polynomials, one for each tract. From the definition of N, O and P it follows that all the f_i's are polynomials in z^{-1} of the order $M_P + M_N + M_O - 6$. For further discussion of the number of poles in the transfer function the reader is referred to page 33.

The numerator of the transfer function (bottom of page 71 and top of page 73) is reordered with the statement:

```
> nHc:=map(factor,collect(Hc*dHc,[O1,O2,O3,O4,N1,N2,N3,N4
]));
```

From the **MapleV**-output the numerator can be rewritten manually:

$$-\tfrac{1}{2} z^{4-M_P/2} \left(\begin{array}{l} (z^2 O_1 - z(\mu_{O1} O_2 - O_3) - \mu_{O1} O_4) k_N z^{-M_N/2} \\ + (z^2 N_1 - z(\mu_{N1} N_2 - N_3) - \mu_{N1} N_4) k_O z^{-M_O/2} \end{array} \right) \tag{A.2}$$

where

$$k_N = (1 + \mu_{PG})(1 + \mu_{N1})(1 + \mu_{NJ}) \tag{A.3}$$

$$k_O = (1 + \mu_{PG})(1 + \mu_{O1})(1 + \mu_{OJ}) \tag{A.4}$$

As the end result of this appendix $H(z)$ can be written as:

$$H(z) = -\tfrac{1}{2} z^{4-M_P/2} \frac{\left(\begin{array}{l} (z^2 O_1 - z(\mu_{O1} O_2 - O_3) - \mu_{O1} O_4) k_N z^{-M_N/2} \\ + (z^2 N_1 - z(\mu_{N1} N_2 - N_3) - \mu_{N1} N_4) k_O z^{-M_O/2} \end{array} \right)}{f_6 z^6 + f_5 z^5 + f_4 z^4 + f_3 z^3 + f_2 z^2 + f_1 z + f_0} \tag{A.5}$$

The order of the transfer function is determined in section 2.5 on page 31.

$$
\begin{aligned}
&- \mu_{pg}\,P4\,N3\,O1 - P2\,\mu_{nj}\,N4\,\mu_{n1}\,O1 + \mu_{pg}\,P3\,O3\,\mu_{nj}\,N3 + \mu_{pg}\,P3\,O3\,N3\\
&+ P1\,\mu_{oj}\,O3\,N4\,\mu_{n1} + P1\,\mu_{oj}\,O4\,\mu_{o1}\,N3 - P2\,\mu_{pj}\,N4\,\mu_{n1}\,O1\\
&- \mu_{pg}\,P4\,\mu_{pj}\,N3\,O1 - \mu_{pg}\,P4\,\mu_{nj}\,N3\,O1 + P2\,\mu_{pj}\,O2\,\mu_{o1}\,N2\,\mu_{n1}\\
&- P1\,\mu_{oj}\,O4\,\mu_{o1}\,N2\,\mu_{n1} + P1\,O4\,\mu_{o1}\,\mu_{nj}\,N3 + P1\,O3\,N4\,\mu_{n1}\\
&- \mu_{pg}\,P3\,\mu_{oj}\,O3\,N2\,\mu_{n1} - P2\,O3\,N2\,\mu_{n1} - P1\,\mu_{nj}\,N4\,\mu_{n1}\,O2\,\mu_{o1}\\
&- P2\,N3\,O2\,\mu_{o1} - N1\,\mu_{pg}\,P4\,O3 - N1\,\mu_{pg}\,P3\,\mu_{oj}\,O4\,\mu_{o1}\\
&+ N1\,\mu_{pg}\,P4\,\mu_{pj}\,O2\,\mu_{o1} - N1\,P2\,O4\,\mu_{o1} - P2\,\mu_{oj}\,O3\,N2\,\mu_{n1}\\
&- \mu_{pg}\,P3\,\mu_{nj}\,N3\,O2\,\mu_{o1} - \mu_{pg}\,P3\,\mu_{nj}\,N4\,\mu_{n1}\,O1 - \mu_{pg}\,P3\,O2\,\mu_{o1}\,N2\,\mu_{n1}\\
&+ P2\,\mu_{pj}\,O3\,N3 + P2\,O3\,\mu_{nj}\,N3 + P1\,O4\,\mu_{o1}\,N3\\
&+ \mu_{pg}\,P4\,\mu_{pj}\,O1\,N2\,\mu_{n1} + P1\,O3\,\mu_{nj}\,N4\,\mu_{n1} + \mu_{pg}\,P3\,\mu_{oj}\,O3\,N3\\
&- P2\,\mu_{pj}\,N3\,O2\,\mu_{o1} - N1\,P2\,\mu_{oj}\,O4\,\mu_{o1}
\end{aligned}
$$

$$
\begin{aligned}
f_4 := &- P1\,O3\,\mu_{nj}\,N3 + P1\,\mu_{nj}\,N4\,\mu_{n1}\,O1 - P1\,\mu_{oj}\,O3\,N3 + N1\,P1\,\mu_{oj}\,O4\,\mu_{o1}\\
&+ P2\,N3\,O1 - P1\,O3\,N3 + P1\,O2\,\mu_{o1}\,N2\,\mu_{n1} - P2\,\mu_{pj}\,O1\,N2\,\mu_{n1}\\
&+ P2\,\mu_{pj}\,N3\,O1 + P2\,\mu_{nj}\,N3\,O1 + \mu_{pg}\,P3\,O1\,N2\,\mu_{n1}\\
&+ P1\,\mu_{oj}\,O3\,N2\,\mu_{n1} + \mu_{pg}\,P3\,\mu_{nj}\,N3\,O1 + N1\,P2\,\mu_{oj}\,O3 + N1\,P2\,O3\\
&+ N1\,\mu_{pg}\,P3\,O2\,\mu_{o1} - N1\,\mu_{pg}\,P4\,\mu_{pj}\,O1 + N1\,\mu_{pg}\,P3\,\mu_{oj}\,O3\\
&+ P1\,\mu_{nj}\,N3\,O2\,\mu_{o1} + N1\,P2\,\mu_{pj}\,O3 - N1\,P2\,\mu_{pj}\,O2\,\mu_{o1}
\end{aligned}
$$

$$
\begin{aligned}
f_5 := &- N1\,\mu_{pg}\,P3\,O1 - N1\,P1\,O2\,\mu_{o1} - P1\,O1\,N2\,\mu_{n1} - P1\,\mu_{nj}\,N3\,O1\\
&+ N1\,P2\,\mu_{pj}\,O1 - N1\,P1\,\mu_{oj}\,O3
\end{aligned}
$$

$$f_6 := N1\,P1\,O1$$

```
> nHc:=map(factor,collect(Hc*dHc, [O1,O2,O3,O4,N1,N2,N3,N4]));
```

$$
\begin{aligned}
\mathrm{nHc} := &\frac{1}{2}\,z^{(6-1/2\,Mp-1/2\,Mn)}\,(1+\mu_{nj})\,(1+\mu_{n1})\,(1+\mu_{pg})\,O1\\
&- \frac{1}{2}\,z^{(5-1/2\,Mp-1/2\,Mn)}\,\mu_{o1}\,(1+\mu_{nj})\,(1+\mu_{n1})\,(1+\mu_{pg})\,O2\\
&+ \frac{1}{2}\,z^{(5-1/2\,Mp-1/2\,Mn)}\,(1+\mu_{nj})\,(1+\mu_{n1})\,(1+\mu_{pg})\,O3\\
&- \frac{1}{2}\,z^{(-1/2\,Mp+4-1/2\,Mn)}\,\mu_{o1}\,(1+\mu_{nj})\,(1+\mu_{n1})\,(1+\mu_{pg})\,O4\\
&+ \frac{1}{2}\,z^{(6-1/2\,Mp-1/2\,Mo)}\,(1+\mu_{oj})\,(1+\mu_{o1})\,(1+\mu_{pg})\,N1\\
&- \frac{1}{2}\,z^{(5-1/2\,Mp-1/2\,Mo)}\,\mu_{n1}\,(1+\mu_{oj})\,(1+\mu_{o1})\,(1+\mu_{pg})\,N2\\
&+ \frac{1}{2}\,z^{(5-1/2\,Mp-1/2\,Mo)}\,(1+\mu_{oj})\,(1+\mu_{o1})\,(1+\mu_{pg})\,N3\\
&- \frac{1}{2}\,z^{(-1/2\,Mp+4-1/2\,Mo)}\,\mu_{n1}\,(1+\mu_{oj})\,(1+\mu_{o1})\,(1+\mu_{pg})\,N4
\end{aligned}
$$

B Glottal signal models

In this appendix two time-domain glottal signal models are reviewed. First the Liljencrants-Fant (LF) model and secondly the Fujisaki-Ljungqvist (FL) model. Various attempts have been made to model the production of the glottal signal e.g. by a two mass model of the vocal cords [Furui, 1989], or a finite element method [Liljencrants, 1991]. The models described in this appendix, however, are curve fitting models made on basis of inverse filtered voiced speech signals.

B.1 The Liljencrants-Fant model

This model was first presented in [Fant et al., 1985] and in short it is often referred to as the LF-model. It has gained wide acceptance and has been used in several research projects [Fant and Lin, 1988], [Fujisaki and Ljungqvist, 1986]. An excerpt from the abstract of [Fant et al., 1985]:

> *"The LF-model is optimal for non-interactive flow parameterization in the sense that it ensures an overall fit to commonly encountered wave shapes with a minimum number of parameters and is flexible in its ability*

to match extreme phonations. Apart from analytically complicated parameter interdependencies, it should lend itself to simple digital implementations."

These properties are exactly the ones desired in this project.

The differentiated airflow is modelled in the time domain thereby including a simple model of the radiation characteristics at the lips and nostrils. Each fundamental period of the glottal signal is expressed in two parts:

$$E(t) = \begin{cases} E_0 e^{\alpha t} \sin\omega_g t & , \ 0 \le t < t_e \\ -\frac{E_e}{\varepsilon t_a}(e^{-\varepsilon(t-t_e)} - e^{-\varepsilon(t_c-t_e)}) & , \ t_e \le t < t_c \end{cases} \quad \text{(B.1)}$$

t lies in the range $[0; t_c]$ where t_c usually is equal to the fundamental period, T_0. Figure B-1 is an example of the flow derivative and the flow in one fundamental period.

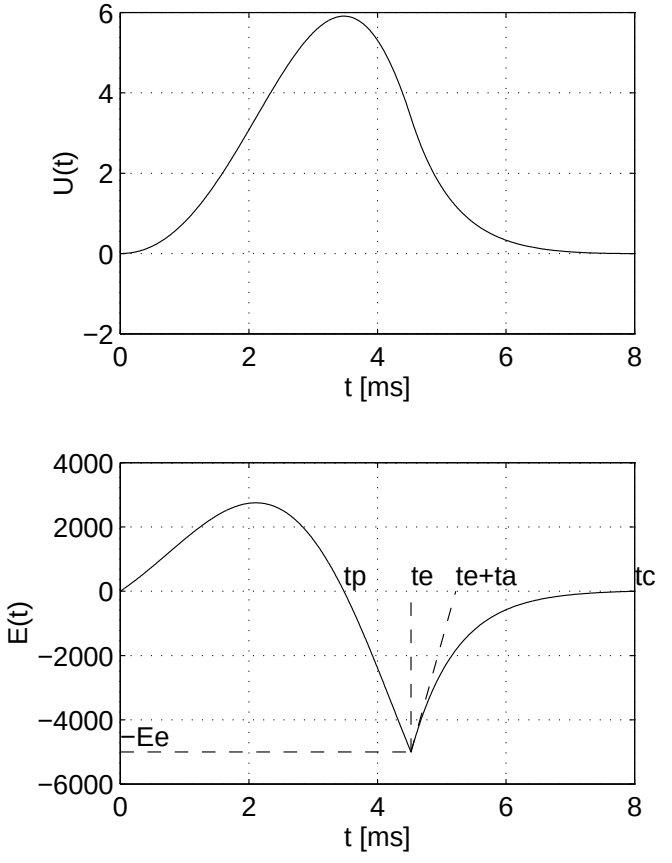


Figure B-1: The Liljencrants-Fant glottal signal model with the glottal flow, *U(t)*, (top) and the differentiated flow, *E(t)*, (bottom). Parameters used: F_0=125Hz, R_a=0.2, R_k=0.3, R_g=1.15, E_e=5000.

The so-called waveshape parameters t_p, t_e, t_a and E_e together with T_0 completely determine the shape of the differentiated flow $E(t)$. Figure B-1 (bottom) illustrates the waveshape parameters. t_a / E_e is the derivative of $E(t)$ at t_e^+. The other parameters in equation (B.1), E_0, α, ω_g and ε, are derived from the waveshape parameters. In this report, however, the normalized parameters, E_e, R_a, R_k, R_g and F_0, are used, see table B-1.

Name	Description
E_e	Maximum negative flow derivative
R_a	The ratio of t_a to $t_c - t_e$
R_k	The ratio of $t_e - t_p$ to t_p
R_g	The ratio of half a fundamental period to t_p

Table B-1: Description of the normalized glottal parameters of the LF-model. All other parameters can be calculated on the basis of these.

For glottal signal models it is common not to count the fundamental period and time offset when stating the number of parameters. Following this custom the LF-model is considered a four-parameter model.

The waveshape parameters are determined as follows

$$T_0 = \frac{1}{F_0} \tag{B.2}$$

$$t_c = T_0 \tag{B.3}$$

$$F_g = F_0 R_g \tag{B.4}$$

$$\omega_g = 2\pi F_g \tag{B.5}$$

$$t_p = \frac{1}{2F_g} \tag{B.6}$$

$$t_e = t_p + R_k t_p \tag{B.7}$$

$$t_a = R_a(t_c - t_e) \tag{B.8}$$

$E_e = -E(t_e)$ and from the second part of (B.1) it is seen that

$$\varepsilon t_a = 1 - e^{-\varepsilon(t_c - t_e)} \tag{B.9}$$

ε can be determined from this equation using numerical root finding methods.

Next step is to determine α, which is accomplished in a few steps. The requirement is imposed, that there may be no buildup of airflow over a fundamental

period. In figure B-1 (top) this corresponds to the requirement that $U(T_0) = U(0)$. This is expressed as

$$\int_0^{T_0} E(t)\,dt = 0 \tag{B.10}$$

or at t_e

$$U_e = \int_0^{t_e} E(t)\,dt = -\int_{t_e}^{T_0} E(t)\,dt \tag{B.11}$$

The right part of this equation can be determined. In [Fant et al., 1985] the integral is approximated by

$$U_e \approx \frac{E_e t_a}{2} K_a \tag{B.12}$$

where

$$K_a = \begin{cases} 2.0 & , R_a < 0.1 \\ 2 - 2.34R_a^2 + 1.34R_a^4 & , 0.1 \le R_a < 0.5 \\ 2.16 - 1.32R_a + 0.64(R_a - 0.5)^2 & , 0.5 \le R_a \end{cases} \tag{B.13}$$

Using the values from figure B-1 this approximation for U_e evaluates to 3.3192. A numerical integration of $-E(t)$ from t_e to t_c yields 3.3805 revealing an error in the approximation of 2% in this case.

If the integral is considered a little closer, however, it turns out that it can indeed be evaluated.

$$\begin{aligned} U_e &= -\int_{t_e}^{T_0} E(t)\,dt = \frac{E_e}{\varepsilon t_a}\int_{t_e}^{t_c} (e^{-\varepsilon(t-t_e)} - e^{-\varepsilon(t_c-t_e)})\,dt \\ &= \frac{E_e}{\varepsilon t_a}\int_{t_e}^{t_c} (e^{-\varepsilon(t-t_e)} + (\varepsilon t_a - 1))\,dt \\ &= \frac{E_e}{\varepsilon t_a}\left[\frac{e^{-\varepsilon(t-t_e)}}{-\varepsilon} + (\varepsilon t_a - 1)t\right]_{t_e}^{t_c} \\ &= \frac{E_e}{\varepsilon t_a}\left(\frac{e^{-\varepsilon(t_c-t_e)} - 1}{-\varepsilon} + (\varepsilon t_a - 1)(t_c - t_e)\right) \\ &= \frac{E_e}{\varepsilon}\left(1 + \frac{\varepsilon t_a - 1}{R_a}\right) \end{aligned} \tag{B.14}$$

As an informal test of equation (B.14) it is evaluated for the values in figure B-1. It yields the correct value of 3.3805. Equation (B.14) is uncomplicated and is more accurate and natural to use than the approximation suggested in [Fant et al., 1985].

The following relation will be used in equation (B.17)

$$-E_e = \lim_{t \to t_e^-} E(t) = E_0 e^{\alpha t_e} \sin\omega_g t_e \Rightarrow \tag{B.15}$$

$$E_0 = \frac{-E_e e^{-\alpha t_e}}{\sin\omega_g t_e} \tag{B.16}$$

The left part of equation (B.11) is considered.

$$\begin{aligned} U_e &= \int_0^{t_e} E(t)\,dt \\ &= \int_0^{t_e} E_0 e^{\alpha t} \sin\omega_g t \;\; dt \\ &= \frac{E_0}{\alpha^2 + \omega_g^2}\left[e^{\alpha t}(\alpha \sin\omega_g t - \omega_g \cos\omega_g t) \right]_0^{t_e} \\ &= \frac{E_0}{\alpha^2 + \omega_g^2}\,(e^{\alpha t_e}(\alpha \sin\omega_g t_e - \omega_g \cos\omega_g t_e) + \omega_g) \\ &= \frac{-E_e}{(\alpha^2 + \omega_g^2)\sin\omega_g t_e}\,(\alpha \sin\omega_g t_e - \omega_g \cos\omega_g t_e + \omega_g e^{-\alpha t_e}) \end{aligned} \tag{B.17}$$

All elements in equation (B.17) are known except α which can be determined using numerical root finding techniques.

B.2 The Fujisaki-Ljungqvist model

The Fujisaki-Ljungqvist (FL) model was first proposed in [Fujisaki and Ljungqvist, 1986] and later used in [Fujisaki and Ljungqvist, 1987]. The model is an attempt to take into account the properties of several previously proposed glottal signal models. Furthermore the computational load of the model is fairly modest, because its is a piecewise polynomial function, whereas most other models use exponential and/or trigonometric functions.

As the Liljencrants-Fant model this is a model of the differentiated airflow. Apart from the fundamental period, T, and time offset the model has six parameters which are given in table B-2.

Name	Description	Value
A	Slope at glottal opening	1000
B	Slope prior to closure	-5000
C	Slope following closure	-3000
D	Glottal closure time	1.5 ms
S	Pulse skew	5
W	Open phase duration	4.5 ms

Table B-2: Parameter description and values used in the example in this section. See also figure B-2 on page 85.

The functional expression of the model is[1]

$$g(t) = \begin{cases} A - \dfrac{2A + R\alpha}{R}t + \dfrac{A + R\alpha}{R^2}t^2 & , \ 0 < t \le R \\ \alpha(t-R) + \dfrac{3B - 2F\alpha}{F^2}(t-R)^2 - \dfrac{2B - F\alpha}{F^3}(t-R)^3 & , \ R < t \le W \\ C - \dfrac{2(C-\beta)}{D}(t-W) + \dfrac{C-\beta}{D^2}(t-W)^2 & , \ W < t \le W + D \\ \beta & , \ W + D < t \le T \end{cases} \quad \text{(B.18)}$$

where

$$W = R + F \quad \text{(B.19)}$$

$$S = \frac{R + F}{R - F} \quad \text{(B.20)}$$

$$\alpha = \frac{4AR + 6FB}{2R^2 - F^2} \quad \text{(B.21)}$$

$$\beta = \frac{CD}{D - 3(T - W)} \quad \text{(B.22)}$$

1. The expressions given in [Fujisaki and Ljungqvist, 1986] and [Fujisaki and Ljungqvist, 1987] both differ slightly from this description - conceivably as results of misprints. The expressions in equations (B.18)-(B.22) are in accordance with the textual descriptions and figures in the articles. Furthermore the requirement of no buildup of airflow over a fundamental period, analogous to equation (B.10), is satisfied.

Examples of the differentiated flow and the flow are shown in figure B-2.

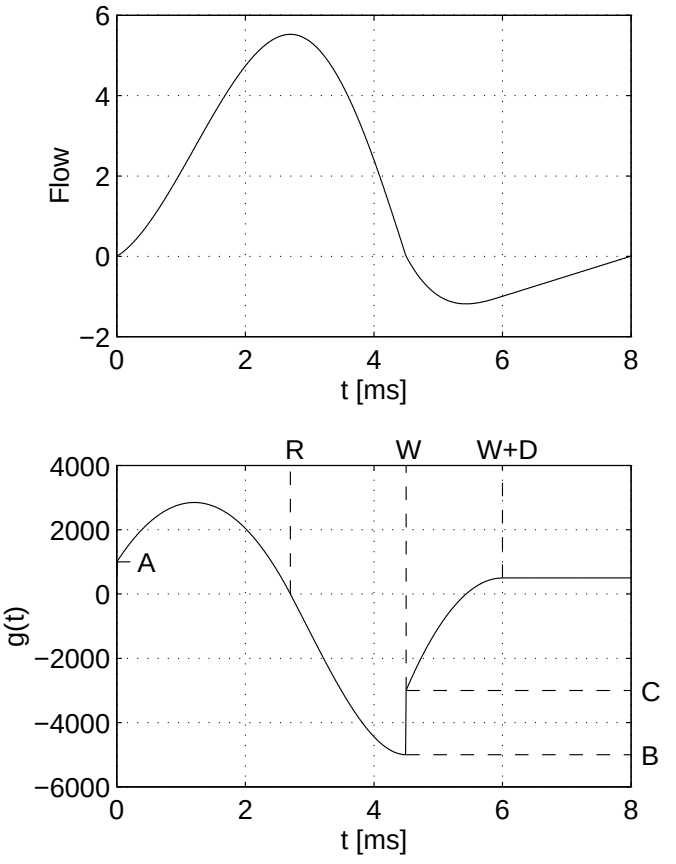


Figure B-2: The Fujisaki-Ljungqvist glottal signal model with the differentiated flow, *g(t)*, (bottom) and the glottal flow (top). The parameter values are taken from table B-2.

Brief tests of the FL-model used in the GARMA analysis described in chapter 3 indicated that a wider variety of glottal waveshapes are possible compared to the LF-model. It seems, however, that this to a large extent is a deficiency of the FL-model, since many of the extra waveshapes were not judged as possible real glottal signals.

C Speech recordings

The purpose of the speech recordings described in this appendix is to obtain a set of files of sampled speech without acoustic reflections and background noise from the room used for the recordings and without phase– and amplitude distortions from the recording equipment. To avoid reflections and background noise the recordings are made in a anechoic room separated from the recording equipment and operators.

Distortions cannot be completely avoided but it is possible to correct the recorded speech in order to compensate for it. The principle of this correction is to determine the transfer function of the system used for the recordings and subsequently process the recorded signals digitally in a way that linearises the phase characteristic and reshapes the amplitude characteristic as desired.

This appendix describes the procedure for the correction, and the outline is as follows: the system used for the recordings is divided into an analog part and a digital part and the transfer functions are determined for both (sections C.3 and C.4 respectively).

The analog part of the transfer function is determined with a measurement. The equipment used for this measurement unfortunately has some effects on the measurement itself. These effects are partly determined by a reference measurement and partly by known characteristics published in a data handbook. Once these undesirable effects have been found the measured analog part of the transfer function is corrected. The reader will benefit from noting the two independent corrections described in this appendix: 1) the correction of the measurement of the analog part of the transfer function (section C.3.2) and 2) the correction of the speech signals (section C.5).

Synchronously with every speech recording a recording is made of the signal from an accelerometer mounted on the speaker's larynx. This recording can be used for pitch-synchronization in the analysis of the speech. The recorded utterances and their corresponding filenames are described in Appendix D on page 103.

C.1 Recording procedure

To manage and log the recording session the EUROPEC[1] software package is used. The software runs on a standardized PC-compatible computer equipped with a plug-in board to sample and process two signals synchronously. The plug-in-board is an OROS-AU22 which has two input channels each with an analog variable-gain amplifier, antialiasing filters and a 16-bit A/D-converter. Two similar output channels are also provided but not used in this case. A TMS320C25 DSP serves as processing unit on the plug-in-board and with the EUROPEC software it is used for highpass and lowpass filtering and up- and downsampling. With the EUROPEC software the system applies oversampling by factors of 2 or 4 where possible. In these recordings the virtual (or desired) sample frequency is set to 32 kHz. In this case the system actually samples at 64 kHz, applies a phase linear digital low pass filter and downsamples by a factor of 2 before storing the signals on files.

The controlling input to the EUROPEC system is a set of files specifying the

- Recording conditions (extension `.rcd`): e.g. virtual sampling frequency, channel amplification and trigger mode.
- General setup (extension `.set`): e.g. pathnames for files used.
- Speaker identification database (`speakers.dbf`): e.g. sex, age and language of speakers.

1. The EUROPEC system was developed as a part of the pan-european Esprit project 2589 called SAM (Speech Assessment Methodologies).

- Prompts (extension `.txt`): the ortographic text corresponding to the utterances in each signal file. The prompts are reproduced in appendix D.
- Corpus (`corpus.dbf`): the set of `.txt`–files in the session.

During the recording session an operator controls the EUROPEC software and the speaker is prompted on a screen with the text corresponding to each signal file. VU-meters with peak detectors are displayed on the computer screen. If mispronunciation or other errors occur it is simple to repeat the recording of the signal file.

Output of the system are files with names like `mome0040.sds` where characters 1 and 2 are the speaker's identification code, characters 3 and 4 identify the `.txt` file and character 5 through 8 constitute a file number unique for the session. For each `.txt` file in the speech corpus four files are produced:

- Speech signal (extension `.sds` or `.pds`).
- Accelerometer signal (extension `.sd2` or `.pd2`).
- Configuration (extension `.cfg`).
- Orthography (extension `.sdo` or `.pdo`): e.g. texts, times and levels.

C.2 Equipment setup

The equipment setup for the recording is shown in figure C-1. The microphone and accelerometer with preamplifiers are located in the anechoic room. Both signals are fed through a measuring amplifier and the microphone signal is low-pass filtered. Finally the signals are fed to the inputs of the OROS board. The apparatus and settings involved is shown in table C-1.

No.	Description	Type	Serial no.	Settings
1	Microphone	Brüel&Kjær 4133	AUC6548	Protection grid removed
2	Micr. preamp	Brüel&Kjær 2619	971106	—
3	Measuring amp	Brüel&Kjær 2636	AUC8717	Input section gain 30 dB Output section gain 0 dB Preamplifier input Linear 1-200000 Hz
4	Filter	Krohn–Hite 3343	AUC8434	Low pass RC Cutoff freq 14000 Hz Gain 0 dB

Table C-1: Apparatus and settings for recording session. Refer to figure C-1.

5	Dual AD/DA PC plug in DSP–board	OROS-AU22	—	Line in 1: microphone Line in 2: accelerometer Gain 1: 12dB Gain 2: 0dB Virtual sampling freq 32kHz
6	Accelerometer	Brüel&Kjær 8307	—	Cable AO0037
7	Vibration pick-up preamp	Brüel&Kjær 2605	AUC7017	X channel Displacement, 300Hz
8	Measuring ampl.	Brüel&Kjær 2636	AUC8022	Input section gain 30 dB Output section gain 0 dB Direct input Linear 1-200000 Hz

Table C-1: Apparatus and settings for recording session. Refer to figure C-1.

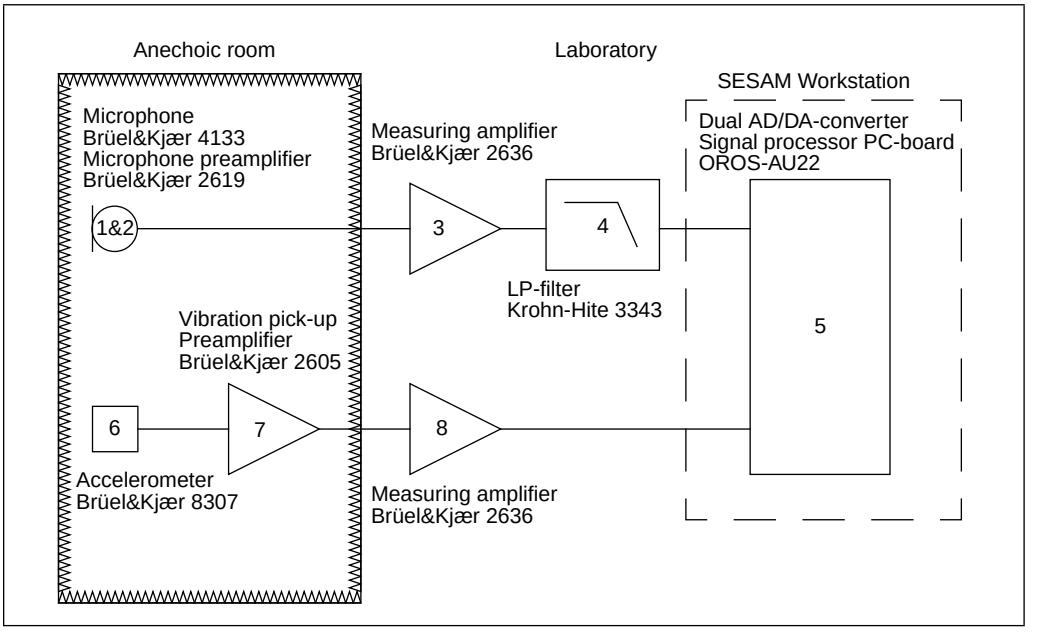


Figure C-1: Equipment used in the speech recordings. Refer to table C-1.

C.3 Analog transfer function

In order to correct the speech signal for any phase- and amplitude distortion that occurs during the recordings, it is essential to know the transfer function from the microphone to the resulting stored speech signal. This transfer function can be divided into an analog part from the microphone to the sample/hold circuits on the OROS-board and a digital part which is the signal processing that takes place in the EUROPEC system by means of the TMS320C25 DSP.

The analog part comprises

- Microphone
- Microphone preamplifier
- Measuring amplifier
- LP-filter
- Two variable-gain amplifiers on the OROS-board
- Analog low pass filter on the OROS board

C.3.1 Measurement of analog transfer function

For the purpose of measuring the analog part of the transfer function, a PC-based Maximum-Length Sequence System Analyser (MLSSA) is used [Rife, 1990]. This system consists of a hardware part and a software part. The hardware is a PC plug-in board and it is used for analog amplification, filtering and sampling an input signal and corresponding generation of an output signal. The software part which runs on the PC is an interactive environment for generation of stimulation signals, acquisition of corresponding system responses and various forms of analysis and display of system properties.

The principle of the MLSSA system is to generate a so-called Maximum-Length pseudorandom stimulus as input to the system for which the transfer function is desired and then measure the response of the system. From the cross correlation between the known stimulus and the measured response it is possible to calculate the impulse response of the system from which the transfer function can be derived by an FFT. The Maximum-Length stimulus has a number of advantageous features:

- The resulting measurements have very high SNR.
- It is periodic which gives a periodic system output. Therefore the system can be allowed to achieve steady-state before the actual measurement and a single period (or any whole number of periods) will contain all information of the system under investigation with no truncation effects.
- Since the stimulus is a calculated function known by the MLSSA system there is no need for measurements of it.
- It permits efficient calculation of the impulse response.

The setup for the measurement is shown in figure C-2 and table C-2 lists the instruments involved and their settings. Units 1 to 5 are unchanged from table C-1.

During normal operation of the electrostatic microphone a thin nickel diaphragm constitutes a capacitance together with the backplate inside the microphone. An acoustic wave results in a movement of the diaphragm and the

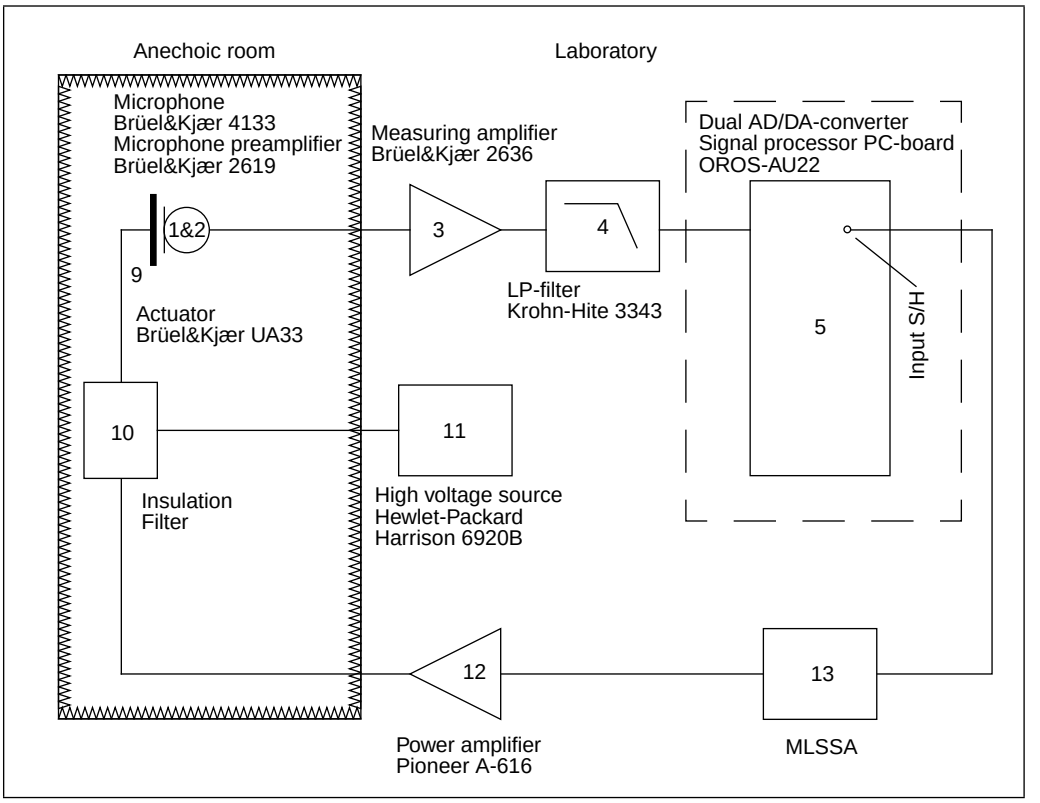


Figure C-2: Setup for the measurement of the analog part of the system transfer function. Refer to table C-2.

No.	Description	Type	Serial no.	Settings
9	Electrostatic actuator	Brüel&Kjær UA33	—	
10	Insulation filter	—	—	
11	High voltage source	Hewlet Packard Harrison 6920B Meter Calibrator	AUC6825	800 volts
12	Power amplifier	Pioneer A-616 Reference Stereo Amplifier	AUC8249	Line input right Speaker A output Volume control -24 dB Direct (no tone control)
13	MLSSA	DRA Laboratories	—	Setup file: `kort2.set` Sample frequency 80kHz Acquisition length 8192 Input gain 10 (±0.5 volts) Stimulus 1.005 volts Stimulus length 65535

Table C-2: Additional apparatus and settings for the measurement of the analog transfer function. Refer to figure C-2.

capacitance is inversely proportional to the distance between the diaphragm and the backplate. As the electric charge of this capacitor is constant in the

audio frequency range, the varying capacitance will result in a varying voltage across the capacitor which is inversely proportional to the capacitance. Consequently the voltage variation is directly proportional to the movement of the diaphragm. A Brüel&Kjær preamplifier serves as an impedance adapter with an input impedance of 10 GΩ and an output impedance of 25 Ω.

If the microphone where activated acoustically by means of a loudspeaker during the MLSSA measurement of the transfer function inevitably the transfer function of the loudspeaker would be contained. To avoid this an electrostatic actuator is applied instead of a loudspeaker for the activation of the microphone. The actuator is a metal grid, which is specifically designed for the half inch Brüel&Kjær microphones, positioned at a fixed distance from the diaphragm and electrically isolated from it. The principle of the actuator is that the diaphragm is moved by means of an electric field of varying strength between the diaphragm itself and the actuator. The varying electric field is generated by a signal added to an 800 volt DC source and led to the actuator grid. An insulation filter shown in figure C-3 serves the purpose of adding the signal to the

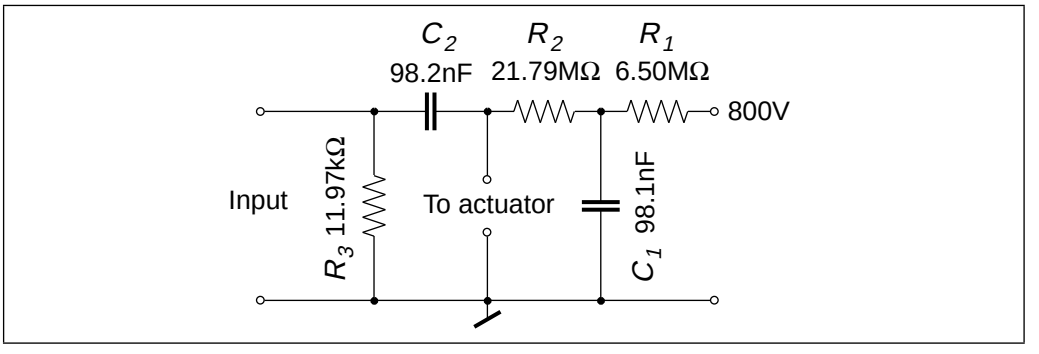


Figure C-3: Diagram of the insulation filter. The component values have been measured.

high voltage DC while protecting the equipment connected to the input from it. Another function of the insulation filter is to low pass filter the high voltage in order to remove noise components from the DC generator.

Figure C-4 shows the effect of the insulation filter as it has been calculated from the measured component values. Frequencies above 0.1 Hz are passed through from the signal input of the filter to the actuator. At frequencies above 40 Hz the phase distortion is less than 0.1˚. From the high voltage input only frequencies near DC are passed to the actuator thereby removing noise components.

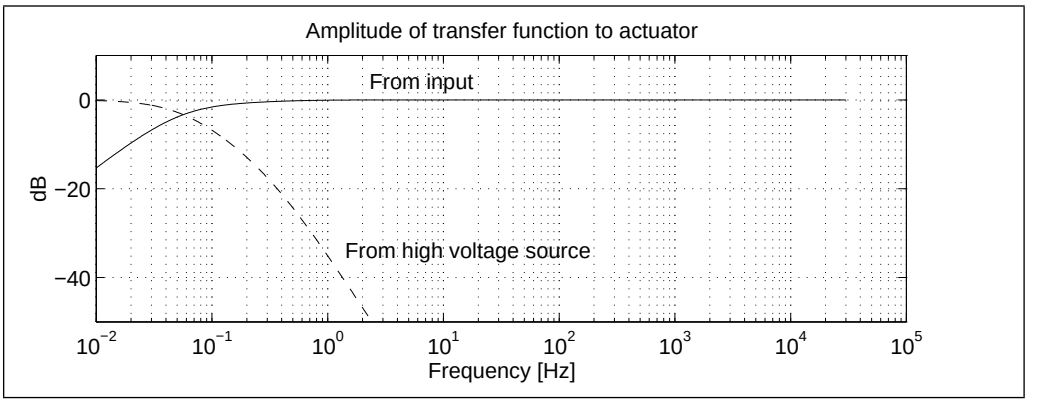


Figure C-4: Transfer function from input of insulation filter to actuator and from high voltage source to actuator.

Among the results of the MLSSA-measurement is a file with the impulse response for the system comprising

- the analog part of the recording equipment
- the characteristics of the actuator (provided in the Brüel&Kjær data handbook)
- the equipment for the measurement

The Fourier transform of this impulse response is the transfer function which is shown in figure C-5 (Measured).

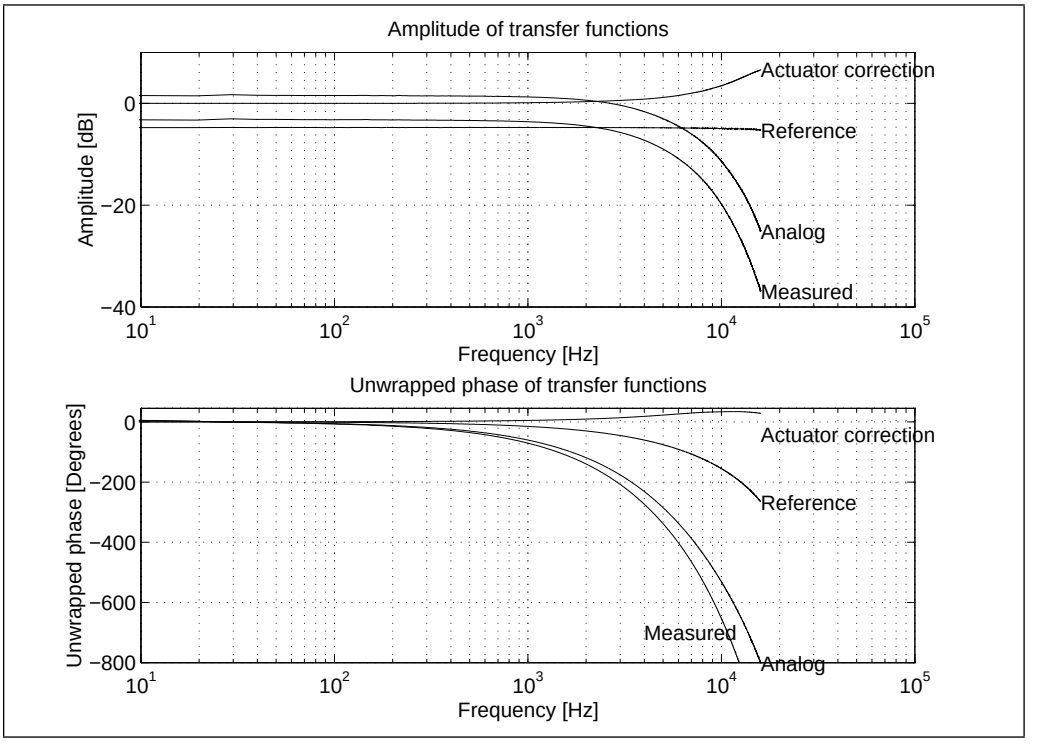


Figure C-5: Transfer functions of the Measured system, Reference system, Actuator correction and the Analog part of the recording system.

C.3.2 Corrections of analog transfer function

It is desirable to remove the contributions from the measurement equipment and to incorporate corrections for the actuator so that the resulting transfer function applies for the acoustic domain.

In order to eliminate the contributions from the measurement equipment it is necessary to identify the transfer function of it and therefore what will be called a reference measurement is carried out. The setup is shown in figure C-6 and table C-3 lists the additional components involved.

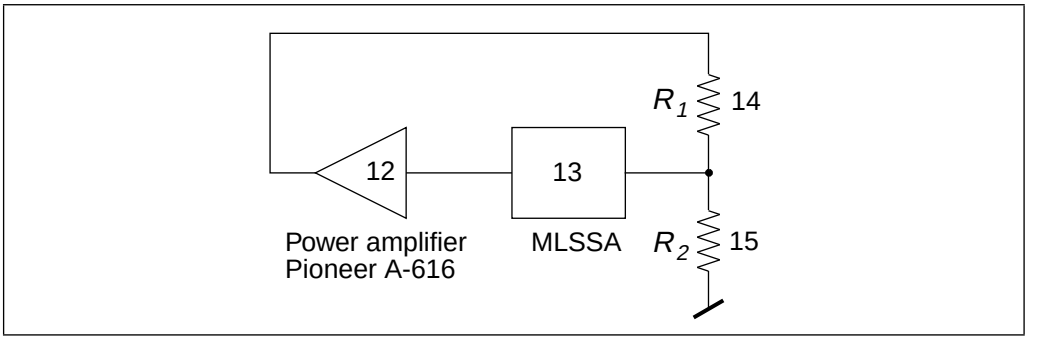


Figure C-6: Setup for the reference measurement.

No.	Description	Value
14	Resistor R_1	1628Ω / 1W
15	Resistor R_2	22.8Ω / 1W

Table C-3: Additional components used in the reference measurement shown in figure C-6.

In the reference measurement the impulse response of the measuring equipment is found. This includes the MLSSA system with filters, power amplifier and cables. Because the actuator operates with an input AC voltage of 30 V and the MLSSA system is set to operate at 0.5 V a simple voltage divider is employed to attenuate the signal. This voltage divider has no effect on the measured transfer function except a frequency invariant attenuation. The insulation filter is not included in the reference measurement, but the contribution from the filter to the transfer function is calculated from the measured component values. The contribution is incorporated in the transfer function of the reference system which is shown in figure C-5 (Reference).

Corrections for the actuator amplitude and phase responses are found in the data handbook for the Brüel&Kjær microphones [Brüel&Kjær, 1982]. These corrections are due to the differences between the influence of the actuator and the acoustic coupling between the air and the microphone diaphragm during normal operation of the microphone, see tables C-4 and C-5.

Freq. [kHz]	0	2	3	4	5	6	7	8	9	10	15	20	30	40	50
Ampl. corr. [dB]	0.0	0.3	0.6	0.8	1.2	1.6	2.0	2.4	3.0	3.5	6.2	7.5	7.6	5.9	5.2

Table C-4: Amplitude correction values for the actuator [Brüel&Kjær, 1982, figure 6.9]. Incidence 0° and protection grid removed.

Freq. [kHz]	0	1	2	5	8	10	12	15	20	25	30	35	40
Phase corr. [^{o}]	0	4.75	9.25	22.5	31.5	33.5	34	30.5	16.5	5.5	2.5	1.75	1

Table C-5: Phase correction values for the actuator [Brüel&Kjær, 1982, figure 6.44]. Incidence 0° and protection grid removed.

These values have been interpolated using a cubic spline algorithm and subsequently plotted in figure C-5.

Taking the complex transfer function corresponding to the measured impulse response, dividing it by the reference measurement transfer function and multiplying it by the actuator correction spectrum the desired analog part of the system transfer function is obtained. This is shown in figure C-5 (Analog). The calculation of the corrected transfer function is carried out in the MATLAB-program in appendix E on page 109.

C.4 Digital processing during recording session

Although the virtual sample frequency is set to 32 kHz the EUROPEC software automatically selects a physical sample frequency of 64 kHz and performs digital low pass filtering, downsampling and high pass filtering. By inspection of the EUROPEC source code for the PC (written in the C language) and for the TMS320C25 processor (PASCAL and assembly languages) it has been established that the antialiasing filters are phase linear FIR filters. These filters can be disregarded because the group delay is without significance and the amplitude alterations are negligible at the frequencies of interest (below 8 kHz).

Another digital signal processing element is a high pass filter for DC and infrasound removal. The following excerpt from the a PASCAL module (`dskop22.inc`) indicates the implementation as a first order IIR high pass filter:

```
DELAY := VIN[I] * COEFF + DELAY * (1 - COEFF)
VOUT[I] := VIN[I] - DELAY
```

The coefficient has the value `COEFF=0.01`. This filter has the transfer function

$$H_{hp}(z) = 0.99\frac{1 - z^{-1}}{1 - 0.99z^{-1}} \tag{C.1}$$

which has a zero at DC and a real pole at 0.99. The pass band amplification is $1.98/1.99 \approx 0.99497$. The transfer function is shown in figure C-7. The cutoff

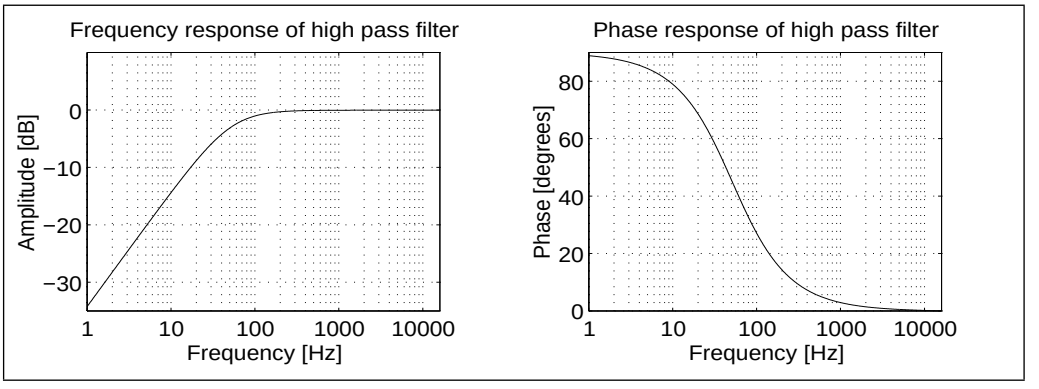


Figure C-7: Transfer function of the original high pass filter.

frequency is 50.7 Hz with $f_s = 32$ kHz but the filter has severe effects in regard of amplitude and especially phase distortions at much higher frequencies. Therefore the high pass filter should be taken into account in the correction of the speech signals.

C.5 Correction of speech signals

In the correction it is desired to linearise the phase characteristics and reshape the amplitude characteristics to an appropriate bandpass function appropriate for downsampling the signal by a factor of two. This is carried out in the following steps

1. Linearisation of the phase characteristics introduced by the analog part of the transfer function by leading the signal in reverse order through a digital filter that imitates the analog part of the transfer function.
2. High pass filtering the signal in reverse order.
3. Removal of the effects of the original digital high pass filter.
4. Shaping of the amplitude characteristics to a low pass characteristic and linearisation in the passband.
5. High pass filtering with the signal in normal order by the filter used in step 2.
6. Downsampling.

C.5.1 Linearisation of the phase characteristics introduced by the analog part of the transfer function

If the original speech signal before analog processing is denoted $s(n)$, the impulse response of the analog system is $h_A(n)$ and the corresponding output signal after analog processing is $s_A(n)$ then

$$s_A(n) = h_A(n) * s(n) \tag{C.2}$$

If the signal $s_A(n)$ is sent through the system in reverse order (last sample first) the output will be

$$s_{Azp}(n) = h_A(n) * s_A(-n) = h_A(-n) * s_A(n) \tag{C.3}$$

The z-transform of $h_A(-n)$ is

$$\begin{aligned} Z(h_A(-n)) &= \sum_{n=-\infty}^{\infty} h_A(-n) z^{-n} = \sum_{n=-\infty}^{\infty} h_A(n) z^{n} = \sum_{n=-\infty}^{\infty} h_A(n) (z^{-1})^{-n} \\ &= \sum_{n=-\infty}^{\infty} h_A(n) (z^{*})^{-n} = H_A^{*}(z) \end{aligned} \tag{C.4}$$

where * denotes the complex conjugate. Consequently

$$S_{Azp}(z) = H_A^{*}(z)\, S_A(z) = H_A^{*}(z)\, H_A(z)\, S(z) \tag{C.5}$$

Since

$$arg(H_A^{*}(z) H_A(z)) = 0 \tag{C.6}$$

and

$$\left| H_A^{*}(z) H_A(z) \right| = \left| H_A(z) \right|^2 \tag{C.7}$$

it is clear that $s_{Azp}(n)$ has the same phase as $s(n)$ ($_{zp}$ is a mnemonic for zero phase) but the amplitude must be corrected for $\left|H_A(z)\right|^2$ in order to eliminate the analog part of the transfer function.

To linearise the phase characteristics of the analog part of the transfer function the procedure is to design a digital filter with the same transfer function (phase and amplitude) as the measured analog one and filter the recorded speech signal in reverse order. The design of the filter is carried out using the truncated impulse response as the coefficients of an FIR filter as shown in the MATLAB-program in appendix E on page 109.

C.5.2 High pass filtering the signal in reverse order

The original digital high pass filter applied by the EUROPEC software during the recordings (section C.4) has undesired effects in terms of phase- and amplitude characteristics and must be replaced by a zero phase (or linear phase) high pass filter. The removal of the effects of the original filter is described in section C.5.5. Rather that applying a linear phase FIR filter a more efficient way in terms of attenuation and computation is to use an IIR filter of comparatively low order and lead the signal through it in both normal (section C.5.5) and reverse directions (this section). As shown in section C.5.1 this gives a zero phase filter with the amplitude characteristic applied twice.

For this purpose a Butterworth characteristic is suitable because of its maximally flat passband and its sufficiently steep cut-off properties. The filter is designed as a second order Butterworth filter with a 3 dB cutoff frequency of 20 Hz in the last part of the MATLAB-program in appendix E. The transfer function resulting from applying this filter twice with the signal in opposite directions is shown in figure C-8.

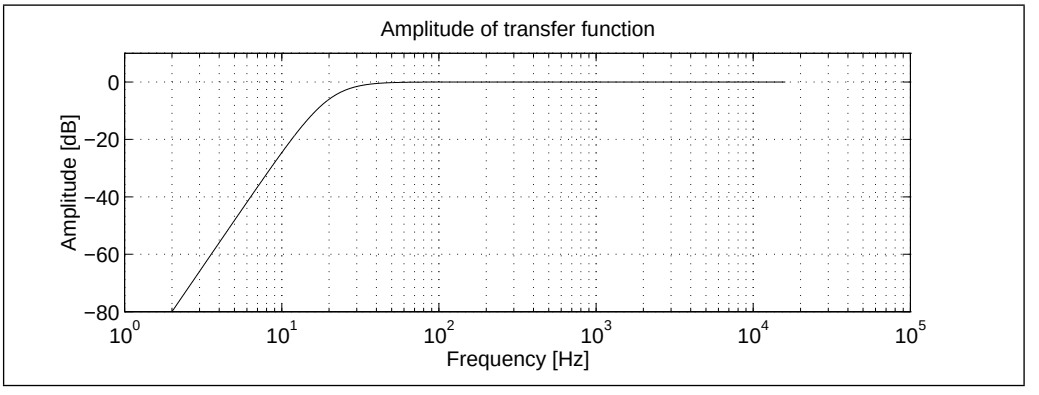


Figure C-8: Transfer function of zero phase high pass filter.

C.5.3 Removal of the effects of the original digital high pass filter

As mentioned in section C.4 the transfer function of the first order IIR highpass filter (equation C.1) contains a zero at DC and a pole at *0.99*. Leading the signal through a filter with a pole at DC and a zero at *0.99* would in principle restore the signal, but this filter would integrate all DC-components including the slightest round off offsets and degrade the accuracy. Therefore a zero is placed at *0.99* and a real pole very close to the unit circle effectively moving the cut-off frequency from 50.7 Hz to 0.1 Hz thereby avoiding DC buildup. The correctional filter has the transfer function

$$H_{chp}(z) = \frac{1+c}{1.98}\,\frac{1-0.99z^{-1}}{1-cz^{-1}} \tag{C.8}$$

resulting in the combined transfer function using equations (C.1) and (C.8)

$$H_{bhp}(z) = H_{hp}(z)H_{chp}(z) = \frac{1+c}{2}\,\frac{1-z^{-1}}{1-cz^{-1}} \tag{C.9}$$

which gives unity passband gain. By setting the absolute value of equation C.9 to $1/\sqrt{2}$ at a frequency of 0.1 Hz and solving for c a solution is found analytically and a value of 0.99998036 can be determined. This filter is realized with the structure in figure C-9. The transfer function of this realisation is given in

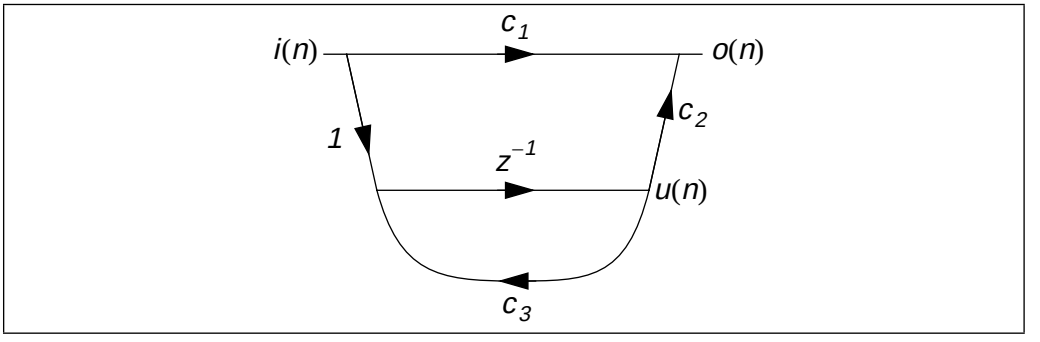


Figure C-9: Structure for realization of filter for removal of original high pass filter effects.

equation (C.10) and the coefficients that match equation (C.8) are listed in table C-6

$$H_r(z) = c_1 + \frac{c_2}{z-c_3} = c_1\frac{1-(c_3-c_2/c_1)z^{-1}}{1-c_3z^{-1}} \tag{C.10}$$

Coefficient	Value
$c_1 = (1+c)/1.98$	1.01009109355489
$c_2 = (c_3-0.99)c_1$	0.01008107803801
$c_3 = c$	0.99998036523867

Table C-6: Coefficient values corresponding to figure C-9.

C.5.4 Shaping of the amplitude characteristics

A linear phase FIR filter is applied with two purposes: 1) linearisation of the amplitude characteristics introduced during the recordings by the analog part of the transfer function and 2) subsequent low pass filtering as preparation for

downsampling. The principle of the design of this filter (appendix E) is the following:

- Establish the desired transfer function on the basis of the corrected MLSSA measurements, which are sampled at 80 kHz.
- Take the inverse DFT of the transfer function to obtain the impulse response.
- The impulse response exhibits high energy segments at both ends. Wrapping the tail to the front (exploiting the periodic property of the DFT) gives a response that is symmetric around zero. Windowing this by a Blackman window and shifting it to become causal gives the 80 kHz filter coefficients. A blackman window is used because of its high side lobe attenuation.
- The coefficients are downsampled to 32 kHz.

In the passband the desired transfer function is the inverse of the square of the analog transfer function previously found (refer to section C.5.1). The low pass characteristic is defined by setting all components of the transfer function above the cutoff frequency (7700 Hz) to zero. Figure C-10 shows the resulting transfer function of the FIR filter. The attenuation above 8 kHz is more than 68 dB.

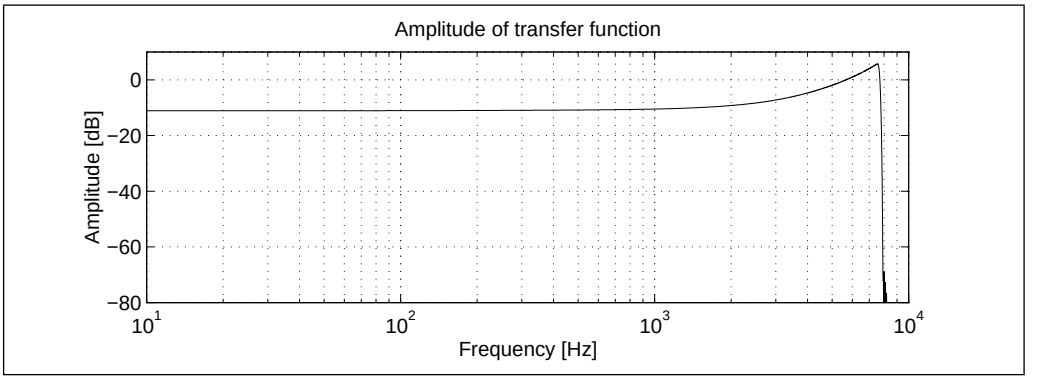


Figure C-10: Transfer function of the FIR filter for shaping of the amplitude characteristics.

C.5.5 High pass filtering with signal in normal order and subsequent downsampling

Again the signal is filtered by the Butterworth filter described in section C.5.2, however this time with the signal in normal order. As mentioned these two passes with the signal in opposite directions effectively constitute a zero phase filter. After all the corrections described in section C.5 have been performed, the phase is linear compared to the original acoustic domain and the amplitude

of the transfer function shown in figure C-11 exhibits a bandpass characteristic appropriate for downsampling which is the final step in the signal processing.

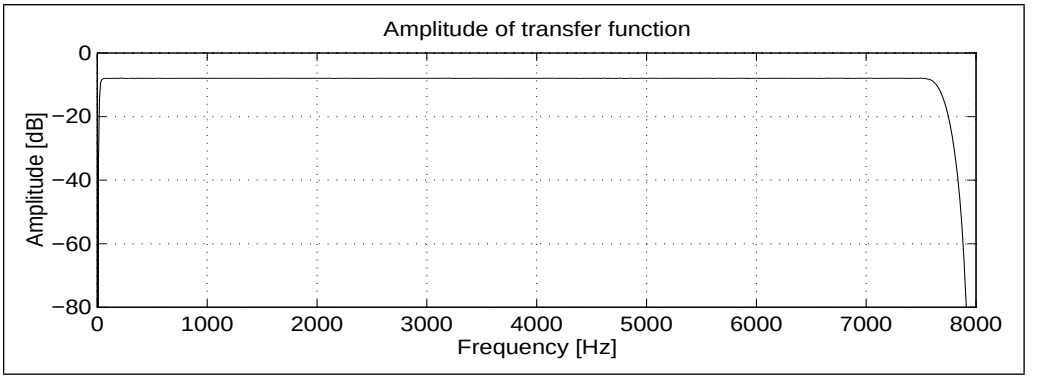


Figure C-11: Resulting transfer function from acoustic domain to corrected signal.

C.5.6 Programs for equalization and downsampling

As mentioned the design of the correction filters are carried out in the Matlab program listed in appendix E. Specialised C-programs were written for the actual signal processing.

D Recorded utterances

In this section the danish utterances recorded as described in section C.1 on page 88 are listed. For every filename the prompt or prompts presented to the speaker are given as individual paragraphs. The utterances where selected in collaboration with a phonetician on basis of phonetic properties.

Filename	Prompt(s)
`mof00025.sdo`	Tøvejret forvandlede hurtigt snefnuggene til grå pytter på pladsen mellem hytten og pigtrådshegnet.
	Dværgpilene og tjørnebuskene langs det faldefærdige markhegn myldrede af kvidrende småfugleyngel.
	Tjeneren fortalte undskyldende, at menukortet desværre kun bød på kylling, klipfisk og kørvelsuppe med æg.

Table D-1: Filenames and recorded utterances (Sheet 1 of 6).

Filename	Prompt(s)
mof10026.sdo	Gertrud Olsens pakkasse indeholdt en rusten mjødøse, en persianer muffe og en antik fonografmaskine. Chefkonsulenten blev vred, da hans grønne MG punkterede på vej hjem fra Bellevue. Jeg hørte på konventet, at den grønlandske rødørn er ukendt for de fleste.
mof20027.sdo	Pøj, hvor de smagte de røde pølser, vi fik ved pølsevognen på trafikpladsen i Tønder. Ifølge landbrugskonsulenten skal man ikke rynke på næsen af gotlandsfårenes kødydelse. Sypigens vulgære knækorte kjole gjorde naturligvis lykke ved prinsernes sidste hofbal.
mof30028.sdo	Høtyvene og alle de nyindkøbte frøpakker flød hulter til bulter i de snavsede vandpytter. Der er flere af hans jødiske bekendte, der stadigvæk kan tale lidt jiddisch til husbehov. Man siger, at madelskere er vilde med så simple ting som hønsekødssuppe og Høng camembert.
mof40029.sdo	Joachim Palms sangrøst har skaffet ham adskillige fine præmier ved mange skolesangdyster. Vølund-vaskemaskinernes nypris er steget vanvittigt i løbet af de sidste fire et halvt år. Vi er omsider nået til det punkt, hvor den konkrete plan nødvendigvis må offentliggøres i radio og TV.
moma0035.sdo	Huset skulle males Turen gik til Råbjerg mile I træet sad en ugle Du må ikke kyle med tingene
momb0036.sdo	Mon det var en skrøne Han fik ømme fødder Hun gav sig til at synge

Table D-1: Filenames and recorded utterances (Continued) (Sheet 2 of 6).

Filename	Prompt(s)
momc0037.sdo	Hun var uanselig
	Hun fik hyacinter
	Hun var uimodståelig
	Hun lyttede til stereo
	Det stiger hende til hovedet
	Hun følte stigen væltede
	Hun fik et stipendium
	Men stipendiet er specielt
	Det gives kun til stoffrie
	Det var en realitet
	Bare det var større
	Hun elskede store stipendier
momd0038.sdo	De var for mange
	Barnet skulle ammes
	De onde magter var skjulte
	Hun plejede at nynne til arbejdet
	Det var en mani for hende
	En funktion med et minimum
	Der var noget indeni
	Omend det var ukendt
mome0040.sdo	Hvor var Inger mon i uge ni i år
	De gik begge i gang da Bodil bad dem om det
	Så sagde Susannes søn osse farvel
	Franz's far sad fire år i fængsel
moq00041.pdo	Vejmeldingen for Storkøbenhavn i dag tirsdag den 28. november. Der er isglatte veje i hele regionen, der er blevet saltet siden klokken 4. Trafikken er tiltagende på alle indfaldsvejene. Der rapporteres om begyndende kødannelse ved Hans Knudsens Plads. Frakørslen på Helsingørmotorvejen ved Jægersborgvej er spærret på grund af et større harmonikasammenstød.
moq10042.pdo	Jeg vil gerne tale med serviceafdelingen. Mit fjernsyn har været til reparation i næsten tre uger nu, og jeg vil vide, hvornår det er færdigt. De hentede det den 13. og lovede, at det skulle være færdigt i løbet af en uge. Jeg er klar over, at der har været problemer med reservedele, men nu har det varet længe nok. Kan jeg få en endelig dato?

Table D-1: Filenames and recorded utterances (Continued) (Sheet 3 of 6).

Filename	Prompt(s)
moq20043.pdo	Jeg er ked af, jeg måtte melde afbud til festen i lørdags. Jeg havde glædet mig til at se jer igen. Men desværre havde jeg et uheld, lige før jeg skulle ud af døren. Jeg skulle ned i kælderen efter noget og gad ikke tænde lyset. På vej op igen snublede jeg på trappen og forstuvede min ankel.
moq30044.pdo	Der er genvej gennem mosen til mit sommerhus. Nogle af de lokale beboere påstår, at det spøger dernede. Der er ikke mange, der er begejstrede for at gå den vej efter mørkets frembrud. Naturligvis tror jeg ikke på sådan noget overtroisk vrøvl. Jeg holder af turen og nyder den maleriske udsigt.
moq40045.pdo	Jeg prøver at komme i kontakt med Jørgen Juliussen i Ribe. Han er flyttet fra Jernbanegade nummer 17 til en anden adresse i Ribe. Kan De oplyse mig om hans nye nummer? Han er flyttet for ca. tre måneder siden. Så vidt jeg ved, har han ikke fået hemmeligt nummer.
moq50046.pdo	Jeg sad på havemuren og kiggede sørgmodigt på vores køkkenhave. Kålen var fuldstændigt gennemhullet af orm. Resten af bedene lignede nærmest et goldt månelandskab. Hvorfor fik vi dog ikke sprøjtet i tide? Jeg havde allermest lyst til at asfaltere det hele én gang for alle.
moq70047.pdo	Natrapport fra vagthavende på station nummer 14. Der har været seks telefonopkald i løbet af vagten. To forsøg på indbrud, tre tilfælde af gadeuorden ved værtshuse og et tilfælde af groft overfald. Med hensyn til gadeuordenen blev én person anbragt i detentionen, de to øvrige blev sendt hjem. Overfaldsmanden blev arresteret klokken 3:38 og han skal fremstilles i grundlovsforhør i morgen.
moq80048.pdo	Om lørdagen elsker jeg at se tipsfodbold i fjernsynet. Min ven er Brøndbytilhænger, men jeg kan bedst lide Lyngby. Når de en enkelt gang spiller mod hinanden, må jeg se kampen hos nogle bekendte, ellers kommer vi bare op at slås. Efter kampen plejer vi at sludre om resultatet over en øl.
moq90049.pdo	Den hurtigste rute til Amager vil være følgende: Kør ad motorvejen til Jægersborg, derefter ad Lyngbyvejen og Nørre Allé. Drej til venstre ved Tagensvej og følg søerne til Søpavillonen. Så til venstre igen, forbi Rådhuspladsen og Tivoli og endelig over Langebro. Det vil tage ca. 40 minutter.

Table D-1: Filenames and recorded utterances (Continued) (Sheet 4 of 6).

Filename	Prompt(s)
mor00017.pdo	Vi har en udmærket sekretær ansat hos os. Desværre har hun sagt op og rejser med udgangen af næste måned. Familien flytter til New Zealand, de skal rejse via Malaysia og Thailand. Vi kommer allesammen til at savne hende. Hun er den type, der altid kan få andre i godt humør.
mor10018.pdo	Jeg hader mandag morgener, især når det regner. Gaderne er fedtede, og jeg er nødt til at gå meget forsigtigt til stationen. Jeg ville meget gerne kunne tage en taxi, men jeg har ikke råd. Min løn er så ringe, at jeg knap nok har penge til sko! Bare jeg ville vinde en million, så kunne jeg købe en bil.
mor20019.pdo	Kan De anbefale en af restauranterne her i nabolaget. Jeg er lige ankommet her i eftermiddags. Jeg er interesseret i noget virkelig eksotisk. En polynesisk eller indonesisk restaurant for eksempel. Det skal helst ikke være udelukkende vegetarisk.
mor30020.pdo	Min kone har et meget kompliceret rejseprogram i næste måned. Kunne De give mig nogle råd om den mest økonomiske løsning. Hun skal til en række møder fra klokken 9 til klokken 13 i Paris, Brügge, Frankfurt, Rom og Hamburg i løbet af fem dage. Kan De finde nogle passende aftenfly og hotelarrangementer? Min kone vil helst undgå store upersonlige hoteller.
mor40021.pdo	Hej, jeg har en skøn ferie her i Lønstrup. Vejret er varmt, solen skinner, og havet er bare ubeskriveligt. I går gik jeg en tur oppe langs klinterne. Det blæste temmelig meget, og jeg var nær blæst ned. Jeg er blevet meget solbrændt, men man kan tydelig se, at jeg spiser for meget is.
mor50022.pdo	Send en ambulance til Jyllandsvej nummer 9 med det samme. Der er en ældre mand, der er faldet i det glatte føre, og han har måske brækket benet. Han har voldsomme smerter. Der er ensretning på Jyllandsvej i øjeblikket på grund af vejarbejde. Det vil være hurtigst at køre ad Parkvej.
mor60023.pdo	Drengen står midt på gårdspladsen og hugger brænde. Hver gang hans økse rammer træstammen, flyver splinterne om ørerne på ham. Smådyr og insekter søger forgæves dækning. Endnu engang har menneskets ubetænksomhed forstyrret naturens gang. Hvornår vil vi dog lære at tage hensyn til andre end os selv?

Table D-1: Filenames and recorded utterances (Continued) (Sheet 5 of 6).

Filename	Prompt(s)
mor80024.pdo	Min veninde skal til lægen i næste uge for at blive vaccineret. Hun skal til det Fjerne østen på studietur og skal derfor vaccineres mod kolera, tyfus, leverbetændelse, polio og stivkrampe. Jeg tror, hun får det temmelig dårligt et par dage. Hun skal have alle vaccinationerne på én gang. Jeg har helt ondt af hende.
moz00030.sdo	Da jeg gik i gymnasiet, spiste jeg altid en stor portion ymer med mysli hver morgen. Bjarne - det er ham den nyrige med klubbens smarteste golftøj og det elendigste golfhandicap.
moz10031.sdo	Monopoltilsynets nye minimalpriser på relækasser og giropapir virker pjattede. Kongens triumftog blev hurtigt afbrudt af fyråb fra demonstranter i grønlige flyjakker.
moz20032.sdo	Den lokalpatriotiske prorektor røbede, at han er tidligere jysk juniormester i langrend. Chefstewardessens smukke opalring funklede om kap med stjernerne på det dybblå himmelhvælv.
moz30033.sdo	Med skælvende hænder og knastør hals fjernede Sonja den utætte toppakning på motorblokken. Folkemasserne blev urolige, da nationalrådet erklærede al nationalpoesi for bandlyst.
moz40034.sdo	Hønsene skreg og vred sig voldsomt for at slippe væk fra genboens nytjærede halvtag. Naboens yngste tøs kylede fnisende mine Lacoste golfsko i svømmepølen med et stort plask.

Table D-1: Filenames and recorded utterances (Continued) (Sheet 6 of 6).

E Equalization of speech recordings

In this appendix the Matlab program used for the design of the correction filters is listed.

```
%-------------------------------------------------------------------------------------------%
% MATLAB - program for calculation of filter coefficients for filters for correction        %
% of speech recordings.                                                                     %
%-------------------------------------------------------------------------------------------%

% Compute actuator correction ---------------------------------------------------------------
% Interpolation of phase correction values for actuator
fp =[0 1 2 5 8 10 12 15 20 25 30 35 40];     % Frequency points [kHz]
p =[0 4.75 9.25 22.5 31.5 33.5 34 30.5 16.5 5.5 2.5 1.75 1];  % Phase correction points (degrees)
fall=0:40/4096:40;                           % Desired frequency points [kHz]
pall=spline(fp,p,fall);                      % 4097 interpolated phase values

% Interpolation of amplitude correction values for actuator
fa =[0 2 3 4 5 6 7 8 9 10 15 20 30 40 50];   % Frequency points [kHz]
a =[0 4.8 8.2 11.8 16.4 21.8 28.0 34.2 41.4 48.7 86.6 105.3 105.9 82.2 73.0];
                                             % Amplitude correction points [mm]
a=a/13.99;                                   % Conversion from mm on enlarged graph to dB
aall=spline(fa,a,fall);                      % 4097 interpolated amplitude values

% Complex transfer function corresponding to actuator correction
act=(10 .^(aall/20)) .* (cos(pall/180*pi)+j*sin(pall/180*pi));% Real and complex values
act(8192:-1:4098)=conj(act(2:4096));         % Symmetry around Nyquist frequency

% MLSSA measurements (sampled at 80 kHz) ----------------------------------------------------
load sysimp                                  % The impulse response of the system measured by MLSSA
sysfft = fft(sysimp);                        % The corresponding transfer function
```

```
load refimp                                 % The impulse response of the reference system

% Calculation of insulation filter transfer function ------------------------------------------
w=2*pi*fall*1000;
c1=98.1e-9;                                 % Measured value
c2=98.2e-9;                                 % Measured value
zc1=[Inf 1./(j*w(2:4097)*c1)];              % Impedance of C1
zc2=[Inf 1./(j*w(2:4097)*c2)];              % Impedance of C2
r1=6.50e6;                                  % Measured value
r2=21.79e6;                                 % Measured value
zx=r2+1 ./(1 ./zc1+1 ./r1);                 % r2+(c1||r1)
insfft=zx./(zc2+zx);                        % From input to actuator
insfft(8192:-1:4098)=conj(insfft(2:4096));  % Complex conjugate symmetry around Nyquist frequency

%-----------------------------------------------------------------------------------------------
% Correction of analog part of transfer function with the reference measurement and the actuator
% transfer function.
reffft = fft(refimp).*insfft';              % The transfer function of reference system
filtfft=sysfft./reffft.*act';               % Correction ...
filtfft(1)=filtfft(2);                      % Discontinuity at DC is disregarded - a high pass
                                            % filter will deal with it

% Due to round-off errors caused by division by small reffft-elements at high frequencies
% filtfft erroneously increase.
num=2180;                                   % This corresponds to 21279 Hz
filtfft2=filtfft;                           % Make a copy
filtfft2(num+1:8193-num)=zeros(1,8193-
2*num);% Frequency components above this freq. are set to zero

% Find the FIR coefficients for the 80 kHz filter ----------------------------------------------
filtimp=real(ifft(filtfft2));               % The corresponding impulse response
cut=160;                                    % Number of coefficients in the 80 kHz filter
imp=filtimp(1:cut)';                        % The truncated impulse response

% Downsample from 80 kHz to 32 kHz
c=80000/32000;                              % Down sampling factor
sinc=zeros(1,cut);                          % The sinc values for each 80 kHz sample
m=0:cut-1;                                  % 80 kHz indexes
filt32=zeros(1,cut/c);                      % 32 kHz samples (dimensioning)
for n=1:cut/c                               % For each 32 kHz sample
 ax=(c*n-c-m')*pi;                          % Argument vector for sinc function
 sinc(ax~=0)=sin(ax(ax~=0))./ax(ax~=0);     % Values of sinc function for each 80 kHz sample
 sinc(ax==0)=ones(1,sum(ax==0));            % If argument is zero then sinc is set to one
 filt32(n)=imp*sinc';                       % The 32 kHz sample is the sum-product
end

% ---------------------------------------------------------------------------------------------
% Compute the filter coefficients for the phase linear FIR filter for the shaping of the amplitude
desfft=abs(1 ./(filtfft.^2));               % Correction for two times the amplitude of the
                                            % analog part of the transfer function
num=max(find(fall<=7.7));                   % Index corresponding to cutoff frequency of 7.7 kHz
desfft(num+1:8193-num)=zeros(1,8193-2*num); % The desired transfer function is actually a
                                            % low pass filter for antialiasing
desimp=real(ifft(desfft));                  % Corresponding impulse response
len=1023;                                   % Uneven length of FIR filter
len2=floor(len/2);                          % Half the length truncated
coeff(1:len2)=desimp(8193-len2:8192);       % First half is the tail of desimp
coeff(len2+1:len)=desimp(1:len2+1);         % Last half is the head. Coeff is now symmetric.
coeff=coeff.*blackman(len)';                % Windowing. These are the 80 kHz coefficients.

% Downsample from 80 kHz to 32 kHz
sinc=zeros(1,len);                          % The sinc values for each 80 kHz sample
coeff32=zeros(1,len/c);                     % 32 kHz samples (dimensioning)
m=0:len-1;                                  % 80 kHz indexes
for n=1:len/c                               % For each 32 kHz sample
 ax=(c*n-c-m')*pi;                          % Argument vector for sinc function
 sinc(ax~=0)=sin(ax(ax~=0))./ax(ax~=0);     % Values of sinc function for each 80 kHz sample
 sinc(ax==0)=ones(1,sum(ax==0));            % If argument is zero then sinc is set to one
 coeff32(n)=coeff*sinc';                    % The 32 kHz sample is the sum-product
end

%-----------------------------------------------------------------------------------------------
% Design a second order Butterworth IIR high-pass filter. The signal is lead through this filter
```

```
% twice: in reverse and forward directions. This way all phase effects are eliminated and the
% amplitude characteristic is applied twice. Attenuation at DC is infinite.
[deshp_b,deshp_a]=butter(2,20/16000,'high');

% Save the results in ascii files ------------------------------------------------------------------
save filt32.mat filt32 /ascii /double
save coeff32.mat coeff32 /ascii /double
save deshp_coeffs.mat deshp_a deshp_b /ascii /double
```

References

For journals the notation is:
Author (Year). Title. *Journal name*, Volume(Number):Start page-End page.

Barry, W. J. and Fourcin, A. J. (1992). Levels of labelling. *Computer Speech and Language*, 6:1–14.

Bothorel et al. (1986). *Cinéradiographie des Voyelles et Consonnes du Français.*. Travaux de L'Institut de Phonétique de Strasbourg.

Brüel&Kjær (1982). *Condenser Microphones and Microphone Preamplifiers for acoustic measurements, Data Handbook*. Brüel&Kjær.

Char, B. W., Geddes, K. O., Gonnet, G. H., Leong, B. L., Monogan, M. B., and Watt, S. M. (1991a). *Maple V Language Reference Manual*. Springer-Verlag, First edition. ISBN 3-540-97622-1.

Char, B. W., Geddes, K. O., Gonnet, G. H., Leong, B. L., Monogan, M. B., and Watt, S. M. (1991b). *Maple V Library Reference Manual*. Springer-Verlag, First edition. ISBN 3-540-97592-6.

Char, B. W., Geddes, K. O., Gonnet, G. H., Leong, B. L., Monogan, M. B., and Watt, S. M. (1992). *Maple V First Leaves: A Tutorial Introduction*. Springer-Verlag. ISBN 3-540-97621-3.

Cohen, M. and Perkell, J. (1986). Palatographic and Acoustic Measurements of the Fricative Consonant Pair /s/ and /s^{v}/. In *Proc. of the 12th International Congress on Acoustics*. Paper A3-5.

Coker, C. (1976). A Model of Articulatory Dynamics and Control. In *Proc. IEEE, 64*, pages 452–460.

Dalsgaard, P. (1992). Phoneme Label Alignment Using Acoustic-Phonetic Features and Gaussian Probability Density Functions. *Computer Speech and Language*, 6:303–329.

Fant, G. (1960). *Acoustic Theory of Speech Production*. Mouton & Co., The Hague.

Fant, G., Liljencrants, J., and Lin, Q. (1985). A Four Parameter Model of Glottal Flow. In *Speech Transmission Laboratory - Quarterly Progress and Status Report (STL-QPSR)*, number 4, pages 1–13, Department of Speech Communication and Music Acoustics, Royal Institute of Technology (KTH), Sweden. ISSN 0280-9850.

Fant, G. and Lin, Q. (1988). Frequency Domain Interpretation and Derivation of Glottal Flow Parameters. In *Speech Transmission Laboratory - Quarterly Progress and Status Report (STL-QPSR)*, number 2-3, pages 1–21, Department of Speech Communication and Music Acoustics, Royal Institute of Technology (KTH), Sweden. ISSN 0280-9850.

Foldvik, A., Husby, O., Kværness, J., Nordli, I., and Rinck, P. (1991). MRI (Magnetic Resonance Imaging) for Filming Articulatory Movements. In *Proceedings of the 12th International Conference of Phonetic Sciences*, pages 34–36.

Foldvik, A. K., Kristiansen, U., Kværness, J., and de Bonnaventure, H. (1993). A Time-Evolving Three-Dimensional Vocal Tract Model by Means of Magnetic Resonance Imaging (MRI). In *EUROSPEECH*, pages 557–558. Updated revision handed out at the conference.

Fujisaki, H. and Ljungqvist, M. (1986). Proposal and Evaluation of Models for the Glottal Source Waveform. In *ICASSP*, pages 1605–1608.

Fujisaki, H. and Ljungqvist, M. (1987). Estimation Of Voice Source and Vocal Tract Parameters Based On ARMA Analysis and A Model For the Glottal Waveform. In *ICASSP*, pages 637–640.

Furui, S. (1989). *Digital Speech Processing, Synthesis, and Recognition*. Marcel Dekker, Inc., New York.

Guérin, B. (1991). The Measurement of the Acoustic Transfer Function and the Area Function of the Vocal Tract: Methods and Limitations. In *Proceedings of the 12th International Conference of Phonetic Sciences*, 171-176.

Kohler, K. (1990). Segmental Reduction in Connected Speech in German: Phonological Facts and Phonetic Explanations. In Hardcastle, W. and Marchal, A., editors, *Speech Production and Speech Modelling*, volume 55 of *NATO ASI Series D*. Kluwer Academic Publishers.

Liljencrants, J. (1991). Numerical simulations of glottal flow. In *EUROSPEECH*, pages 255–258.

Ljung, L. (1987). *System identification Theory for the user*. Prentice-Hall. ISBN 0-13-881640-9.

Maëda, S. (1982). A Digital Simulation Method of the Vocal-Tract System. *Speech Communication*, 1(3-4):199–229.

Markel, J. and Gray, A. (1976). *Linear Prediction of Speech*. Springer-Verlag.

Nadler, R., Abbs, J., and Fujimura, O. (1987). Speech Movement Research Using the New X-Ray Microbeam System. In *Proceedings of the 11th International Congress on Phonetic Sciences*. Paper Se 11.4.

Olesen, M. (1993). Derivation of the Transfer Function for a Speech Production Model Including the Nasal Cavity. In *EUROSPEECH*, pages 549–552. Paper 16-3.

Oppenheim, A. V. and Schafer, R. W. (1975). *Digital Signal Processing*. Prentice-Hall.

Parthasarathy, S. and Coker, C. (1990). Phoneme-Level Parameterization of Speech Using an Articulatory Model. In *ICASSP*, pages 337–340.

Perkell, J. and Cohen, M. (1986). An Alternating Magnetic Field System for Tracking Multiple Speech Articulatory Movement in the Midsaggital Plane. Technical Report 512, Res. Lab. Electronics, MIT, Cambridge.

Perrier, P. and Boë, L. J. (1989). Passage de la Coupe Saggitale à la Fonction D'Aire: Les Zones de Faibles Dimensions. *Journal Acoustique*, (2):59–67.

Rabiner, L. and Schafer, R. (1978). *Digital Processing of Speech Signals*. Prentice Hall, New Jersey.

Rife, D. D. (1990). *MLSSA (Maximum-Length Sequence System Analyzer), Reference Manual v. 6.0*. DRA Laboratories.

Schroeder, M. R. (1967). Determination of the Geometry of the Human Vocal Tract by Acoustic Measurements. *Journal of the Acoustical Society of America*, 41(4):1002–1010.

Schroeter, J. and Sondhi, M. M. (1994). Techniques for Estimating Vocal-Tract Shapes from the Speech Signal. *IEEE Transactions on Speech and Audio Processing*, 2(1):133–150. Part II.

Scully, C. (1987). Linguistic Units and Units of Speech Production. *Speech Communication*, 6(2):77–142.

Wang, R., Guan, Q., and Fujisaki, H. (1990). A method for robust GARMA analysis of speech. In *ICSLP*, pages 33–36.

Wolfram, S. (1991). *Mathematica, A System for Doing Mathematics by Computer*. Addison-Wesley Publishing Company, Inc., Second edition. ISBN 0-201-51507-5.