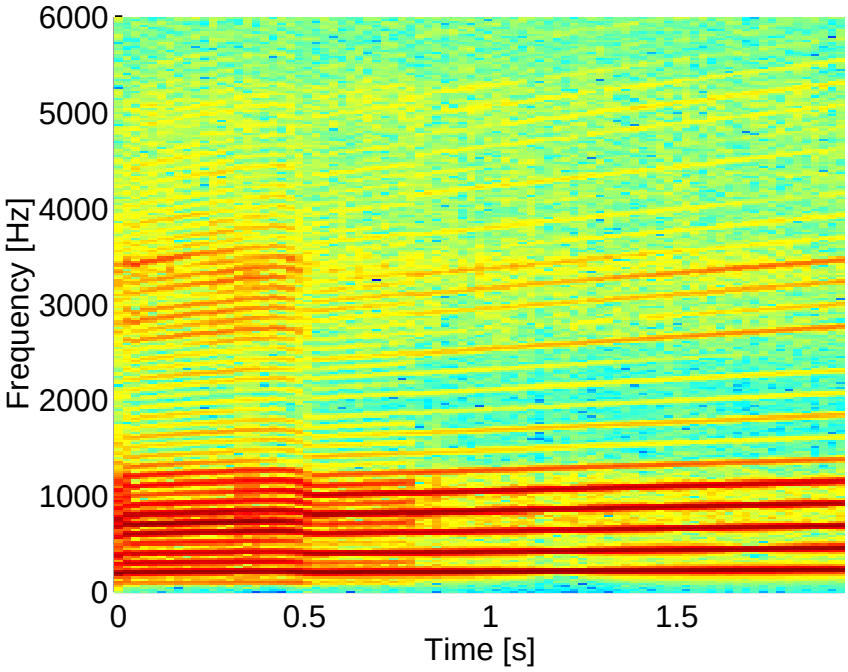
Malte Kob

Physical Modeling of the Singing Voice



PHYSICAL MODELING OF THE SINGING VOICE

Von der Fakulät für Elektrotechnik und Informationstechnik der
Rheinisch-Westfälischen Technischen Hochschule Aachen
zur Erlangung des akademischen Grades eines
DOKTORS DER INGENIEURWISSENSCHAFTEN
genehmigte Dissertation

vorgelegt von

Diplom-Ingenieur
Malte Kob
aus Hamburg

Berichter: Universitätsprofessor Dr. rer. nat. Michael Vorländer
Universitätsprofessor Dr.-Ing. Peter Vary
Professor Dr.-Ing. Jürgen Meyer

Tag der mündlichen Prüfung: 18. Juni 2002

Die Deutsche Bibliothek – CIP-Einheitsaufnahme

Kob, Malte:
Physical modeling of the singing voice / vorgelegt von Malte Kob. - Berlin : Logos-Verl., 2002

Zugl.: Aachen, Techn. Hochsch., Diss., 2002
ISBN 3-89722-997-8

ISBN 3-89722-997-8

Logos Verlag Berlin
Comeniushof, Gubener Str. 47,
10243 Berlin
Tel.: +49 030 42 85 10 90
Fax: +49 030 42 85 10 92
INTERNET: http://www.logos-verlag.de

Meinen Eltern.

Contents

Abstract – Zusammenfassung

Abstract

This thesis deals with the physical modeling of the parts of the voice organ relevant for voice generation plus techniques for the measurement of acoustic voice properties. An introduction to characteristics of the voice signal is followed by a literature survey of existing approaches for the most important functional voice components. Algorithms that seem to be suitable for modeling of the singing voice are adopted and extended.

The modeling of the vocal fold movement uses a three-dimensional, symmetric multiple mass model that is capable of simulating different voice registers and voice pathologies that are found in singers. For the wave propagation in the space between glottis and mouth opening, the vocal tract, two algorithms are presented, which have been optimised for different applications. The first model is based on cylinder segments and requires a fixed sampling rate that yields a high resolution in space. The second model allows an arbitrary choice of the sampling rate and makes it possible to reduce the number of parameters for the description of the vocal tract by using conical segments. Since the noise component is required for a natural sounding voice, a model is implemented that simulates vortex shedding and sound generation by turbulences. The dependence of the noise component on the choice of the articulated speech sound is described by analysis of the voice signal in the domains of time and frequency.

The resonance characteristics of the vocal tract are evaluated with two measurement approaches: a direct method that determines the transfer function and a mobile, non-invasive set-up for the measurement of the acoustic impedance at the mouth.

For comparison of the characteristic radiation of the human voice with an artificial singer, a measurement set-up is described that allows a detailed visualization of the directivity.

The final part of this work investigates the interaction of the elements of the model. Some examples for the application of the singing voice model to the simulation of different singing styles and voice pathologies are presented. Different voice registers are modelled with special emphasis on the simulation of overtone singing. The impedance measurements were the basis for the parameter choice of the vocal tract model.

As a future application of the model, the investigation of voice pathologies is planned. First attempts to model edema of the vocal folds and singer's nodules are presented and the use of the model as a therapeutic tool for voice therapy is discussed.

Zusammenfassung

Diese Arbeit beschreibt die physikalische Modellierung der für die Erzeugung der Singstimme relevanten Stimmorgane sowie Messverfahren zur Bestimmung akustischer Eigenschaften der Stimme. Nach einer Einführung in die Charakteristika des Stimmsignals wird ein überblick über bestehende Modellierungsansätze zur Beschreibung der wichtigsten Stimmfunktionskomponenten gegeben sowie jeweils eine Anpassung und Erweiterung geeigneter Algorithmen für die Modellierung der Singstimme vorgenommen.

Die Modellierung der Stimmlippenbewegung erfolgt mit einem dreidimensionalen, symmetrischen Mehrmassenmodell, das sowohl für die Nachbildung verschiedener Stimmregister als auch zur Simulation von sängertypischen Stimmerkrankungen geeignet ist. Für die Modellierung der Schallausbreitung im Raum zwischen Glottis und Mundöffnung, dem Ansatzrohr, werden zwei Algorithmen vorgestellt, die für jeweils unterschiedliche Anwendungszwecke optimiert wurden. Während beim ersten Modell bei fester Abtastrate eine räumlich hochaufgelöste Diskretisierung anhand von Zylindersegmenten erfolgt, erlaubt das zweite Modell eine freie Wahl der Abtastrate und Reduzierung der für die Ansatzrohrbeschreibung nötigen Parameter durch Verwendung konischer Segmente. Für einen natürlichen Stimmklang ist ein Rauschanteil erforderlich, für den ein Modell verwendet wird, das die Wirbelbildung und Schallerzeugung durch Turbulenzen nachbildet. Die Abhängigkeit des Rauschanteils von der Wahl des artikulierten Phonems wird durch Analyse des Stimmsignals im Zeit- und Frequenzbereich beschrieben.

Zur messtechnischen Bestimmung der Resonanzeigenschaften des Ansatzrohres werden zwei Verfahren vorgestellt: eine direkte Methode zur Messung der übertragungsfunktion sowie ein mobiler, nicht invasiver Messaufbau zur Ermittlung der akustischen Impedanz am Mund.

Für den Vergleich der charakteristischen Abstrahlung der menschlichen Singstimme mit der eines künstlichen Sängers wird ein Messverfahren beschrieben, das eine detaillierte Darstellung der Richtcharakteristik erlaubt.

Im letzten Teil der Arbeit wird die Interaktion der Modellkomponenten untersucht und anhand einiger Beispiele die Anwendung des Singstimmenmodells für die Nachbildung verschiedener Singstile sowie von Stimmstörungen vorgestellt. Neben der Modellierung verschiedener Stimmregister wird insbesondere die Nachbildung von Obertongesang untersucht, wobei die beschriebene Impedanzmessmethode wichtige Hinweise für die Parametrisierung des Ansatzrohrmodells lieferte.

Als Ausblick auf zukünftige Anwendungen des Modells für die Untersuchung von Stimmerkrankungen werden erste Ansätze zur Modellierung von Stimmlippenschwellungen und Sängerknötchen vorgestellt sowie die Verwendung des Modells als didaktisches Werkzeug für die Stimmtherapie diskutiert.

Keywords

physical model, vocal folds, vocal tract, aspiration noise, artificial singer, impedance measurement, pathologic voice, overtone singing

Introduction

> „Ein scharfes Wort soll gesprochen werden über die unbefugte Jagd der stimminteressierten Physiologen bzw. Laryngologen auf die Bestimmbarkeit der Gattung wie der Facheinteilung der Sängerstimmen. Die Bestrebungen der Kehlkopfspezialisten, aus der sichtbaren Bauart der Kehlkopfteile und des Ansatzrohrs (insbesondere des Gaumens) unmittelbar auf das Hörbare, auf Charakter und Klang schließen zu wollen, sind zwar dem Problem nach verständlich. ... Die Forschungen sind noch in keiner Weise abgeschlossen und spruchreif. Und solange sie keine wissenschaftliche Beweiskraft haben, ist in all solchen Fällen der Sänger nichts als Freiwild für tatenlustige Sonntagsjäger.“ [1]

Voice research has significantly intensified in the last decades. Numerous new applications of computer speech, transmission and coding of speech, or voice recognition have found a way into our daily life. Physical models for voice synthesis have often been rejected because of their poor quality or the computational costs. However, high-quality models are increasingly employed as the rapidly growing speed of computers allows researchers to take into account more and more details of voice physiology. A satisfactory on-line modeling of speech based upon the physiology of an individual speaker is not yet possible, even though the advantages of such an approach are tempting:

- the synthesis of individual voices could be based upon the actual physiology
- speech transmission could use a set of physically-based parameters for increased naturalness of the speech
- a comparison of recorded voices with results from simulations could give certain hints for diagnosis

Some of these applications are not available today but are currently developed with efforts in industry as well as at universities. Before such goals can be achieved, basic research is necessary at several stages: measurements of anatomical static and dynamic parameters, mapping of such parameters to mathematical models, development

[1] From: Franziska Martienssen-Lohmann: Der wissende Sänger, Gesangslexikon in Skizzen, Atlantis Musikbuch-Verlag, Zürich, Mainz, 1956.

and implementation of such models, and development of measurement methods for validation of the results.

Physical modeling is the description of observed phenomena by means of mathematical terms. The model will always be an approximation of reality. The challenge for an engineer is to find a compromise between the conformity of the calculated result with reality and the complexity of the model.

The focus of this thesis is laid upon the synthesis of vowels rather than on speech synthesis. The problem of speech generation is a very complex topic that extends from physics via phonetics to linguistics. However, the acoustic part is quite similar to singing voice generation if plosives and fricatives are excluded. In other words, this work deals with the analysis and physical modeling of voiced sounds like vowels.

The thesis is divided into seven chapters. In chapter 1 the signal properties of the singing voice are described and the variety of different singing styles is presented. The chapter concludes with a description of the radiation characteristics of a singer. Physical modeling of the voice can be divided into several functional components. The generator for all voiced sounds are the vocal folds. In chapter 2 vocal fold physiology is briefly described. A new model based upon a description of existing vocal fold models is presented. Parameters of the model are discussed and compared to literature and measurements. The production of different vowels as well as of consonants is achieved by modification of the anatomical space between the vocal folds and the lips. This space is called the vocal tract (in German: Ansatzrohr). Chapter 3 deals with the structure and parametrisation of the vocal tract. Existing models are presented and the implementation of two models is described. At the end of the chapter results from simulations are compared to measured transfer functions and vocal tract mouth impedances for selected phonemes. Chapter 4 investigates the non-harmonic part of the voice signal: the noise. Two kinds of noise sources are described and an algorithm for aspiration noise, i. e. noise generation at the glottis, is presented. The results from simulations are compared to measurements and literature. In chapter 5 the radiation characteristics of a singer are analysed. Measurements of the directivity of human singers and an artificial singer are presented. An outline of the possible interaction between vocal tract and vocal folds, and the applications of the combined model to the synthesis of the singing voice are presented in chapter 6. Emphasis is laid on the generation of overtone singing and the generation of pathological voices. In the last chapter this work is summarised and discussed and applications for educational and medical use are proposed.

Terms in Latin language – such as medical terms – are printed in *italics* (cf. appendix, Table A). The appendix also contains lists of speech sounds, abbreviations, Figures and Tables.

Chapter 1

The singer

Music has most probably been founded by early mankind, when people tried to entertain each other by producing composed sounds. Their voices might have been one of their instruments being available and versatile. The range of sounds that can be produced are as different as the simple voice of an untrained child, the mighty voice of an opera singer, or the magic biphonic sound of an overtone singer. With respect to timbral range and flexibility only very complex musical instruments like organs or synthesisers can compete with the human voice.

This chapter will illuminate the character of the "musical instrument" and sound source *the singer*. The aspects presented in section 1.1 are the perceptual and functional properties of the voice signal and a glancing view on the sound of a choir. A description of different singing styles is given in section 1.2. Due to their complexity, the multitude of aspects of spoken language synthesis is not considered here. A comprehensive study of speech models and data has been published by B. Kröger [Krö98]. Extreme singing styles like growl (Louis Armstrong) will not be considered here, either.

In Figure 1.1 a schematic overview of the integration of the functional components influencing the human regulatory system is given.

In the drawing, the entities body/man (outermost box), brain (middle box) and conciousness/higher-order centers (innermost box) are described with their functions (boxes), flow of information (solid lines) and interactions (dashed lines). The voice generation is initiated by the higher-order centers ('give me the phoneme [a:]!'). The motor cortex forms the necessary laryngeal and articulatory targets for the motor nuclei control of the muscles of the phonatory organs. Muscle activity will set the boundary conditions, i. e. the physical properties of the voice organs. As a result, voice is produced. Of course, the process is much more complex as the interaction between the different functional components continuously changes these boundary conditions, partly intended by the singer, partly without intention. Two levels of control of the phonatory organs can be found: the feedback of sensory receptors at

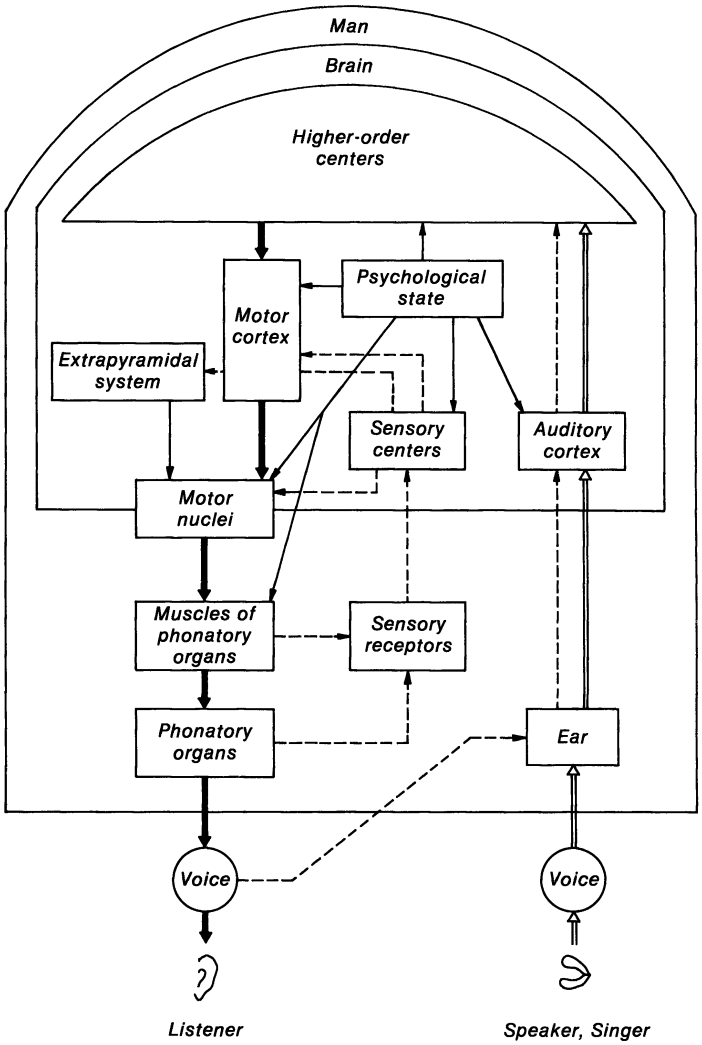


Figure 1.1: Schematic voice production of a singer (after [Hir68])

the muscles of the voice organs to the sensory centers of the brain, and the feedback of the ear to the auditory cortex of the brain. Both feedback loops enable the singer to adjust voice organ parameters like tension of the folds or positioning of the tongue in order to achieve a desired sound quality of the voice.

1.1 Voice signal

An important feature of a musical sound is its variation over time. A categorisation of variation can be made either from the musician's point of view: intended vs. unintended sound variations, or from a perceptual point of view: modulation of signal parameters like amplitude, pitch or timbre.

The singing voice has most features of a musical instrument with a sustained sound: equally spaced harmonics, a rather wide frequency range of the fundamental, and a modulation of the stationary sound. If the singer illustrates the musical content with text, all of these properties plus spectral envelope vary quickly over a huge range of parameters. In the following, these well-known acoustically relevant cues will be discussed.

1.1.1 Harmonic structure

The main character of the sustained voiced sound is the structure of the spectral envelope which contains phonetic information and the timbre.

The harmonics are the peaks of equally spaced frequencies within a spectrum, and the envelope is their amplitude distribution vs. frequency as shown in Figure 1.2.

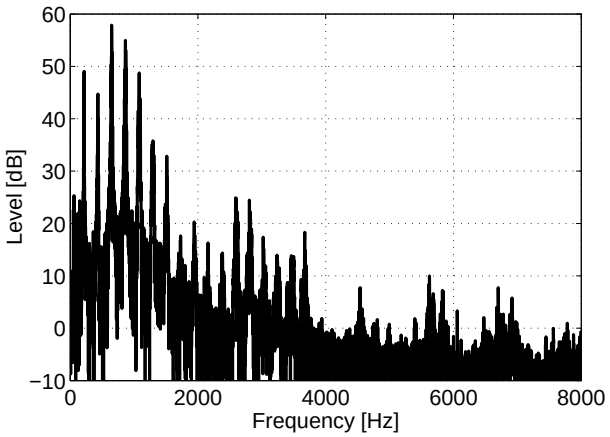


Figure 1.2: Spectrum of a voice signal

In case of normal phonation, the fundamental frequency f_0 determines the lowest frequency of the harmonics. The upper harmonics are integer multiples of f_0, i. e. $2f_0$, $3f_0$ etc. and are well above the noise floor for frequencies below 7 kHz but at least 40 dB, corresponding to a factor of $\frac{1}{10000}$ in power or $\frac{1}{100}$ in amplitude, weaker than the maximum harmonic for frequencies below 4 kHz. If the singer uses vocal fry register (also called pulse, glottal fry, creak or Strohbaß register, see section 1.2.1) or vocal-ventricular mode (VVM)[1] techniques [Fuk99], the lowest harmonic can be an integer fraction of f_0 i. e. $\frac{f_0}{2}$, $\frac{f_0}{3}$ etc. Subharmonics and irregular spacing of the frequency peaks also occur if non-linear effects determine the fold oscillation. These effects will be briefly discussed in chapter 2.

The envelope of the harmonics is determined by the source spectrum of the voice signal at the glottis and the shape of the space between glottis and lips, generally called the vocal tract (VT). The singer's choice of phoneme determines the constrictions within the VT which form an acoustic filter that amplifies and attenuates at certain frequencies. Peaks within the envelope spectrum of such a filtered signal are called formants. The frequency, bandwidth and amplitude distribution of the formants characterise the sound and are specific for each vowel and unvoiced sound.

[1] A list of abbreviations can be found in appendix A.

Apart from the stationary features of the signal, the dynamic aspects have to be considered for speech signals. These aspects are beyond the scope of this thesis.

1.1.2 Pitch and amplitude

Maria Callas, a famous mezzo soprano singer, claimed to have a vocal range, i. e. the difference between lowest and highest note[2], from $F_3^{\#}$ (∼185 Hz) to E_6 (∼1319 Hz) [Fis93]. Almost three octaves is a very big range for musical use for a mezzo soprano voice. The classification of singers is done into the main groups Soprano, Mezzo-Soprano, Alto, Tenor, Baritone, and Basso. In Table 1.1 the range of voices for these groups is shown [Fis93]. The variation of loudness of the singing voice can be

Table 1.1: Ambitus and frequency range of voice groups (after [Fis93]).

	Lower limit		Upper limit	
Group	Note	f_0 [Hz]	Note	f_0 [Hz]
Soprano	$G_3..C_4$	196..262	$D_6..G_6$	1175..1568
Mezzo-Soprano	$F_3..F_3^{\sharp}$	175..185	C_6	1047
Alto	$C_3..E_3$	131..165	$G_5..B_5$	784..988
Tenor	$F_2^{\sharp}..C_3$	92..131	$C_5..G_5$	523..784
Baritone	$F_2..F_2^{\sharp}$	87..92	$G_4..B_4^{\flat}$	392..466
Basso	$C_2..E_2^{\flat}$	65..78	$F_4..G_4$	349..392

expressed by the sound pressure level (SPL). The SPL variation between piano and forte phonation, measured at the mouth, is about 57..92 dB for non-professional male singers and about 60..90 dB for non-professional female singers [NR00]. The standard deviation for each of the groups (10 healthy subjects each) is about 5 dB. However, professional singers achieve much higher SPL values.

Vibrato

Vibrato is an intentional variation of amplitude and fundamental frequency of a sustained vowel. It is used by experienced singers to increase the subjective intensity of a tone without increasing the SPL. Some effort has been made by P.-M. Fischer to describe accurately the characteristics of vibrato [Fis93]. The approaches aim at either describing the singing technique for musical use or analyzing the physical cause of vibrato generation. Most authors cited by Fischer agree on an optimum vibrato frequency to be between 5.5 Hz and 6.5 Hz, while the amplitude variation should not

[2]The musical notation follows the American style. For German notation, the following conversion (American ↔ German) applies: ... $A_1 \leftrightarrow A_1$, $A_2 \leftrightarrow A$, $A_3 \leftrightarrow a$, $A_4 \leftrightarrow a^1$, $A_5 \leftrightarrow a^2$ ...

exceed 3 dB, and the frequency modulation should be about ±1 semitone. However, the underlying process for the generation of vibrato is not clearly understood, yet.

Fischer distinguishes between two origins of vibrato: the respiratory wave (German: Atemwelle) and the glottis wave. The respiratory wave is caused by a rhythmic movement of the *diaphragma* and the *peritoneum* with a frequency of 3.5 Hz to 4 Hz that causes a frequency modulation of ±1 semitone. The glottis wave is caused by a rhythmic contraction of the glottis muscles with a frequency of 6.5 Hz to 8 Hz that varies individually in amplitude compass of voice and frequency modulation. The optimum vibrato should combine both origins resulting in the desired complex vibrato.

Jitter and shimmer

Jitter and shimmer are expressions for small variations in period length and amplitude of a signal. With respect to these parameters, the voice of a singer is far from being a stationary sound. Small fluctuations are perceptually important for the naturalness of a voice sound. If jitter and shimmer are absent, the voice sounds clean and boring, in other words: synthetic. Therefore, algorithms for the additive synthesis of natural sounding musical signals always include a residual component, often white noise [Goo96]. However, care must be taken to synchronise the harmonic and residual part of the signals [Her91].

The origin of jitter and shimmer can be found in the function of vocal fold muscles as well as in the variations of the vocal tract geometry. A comprehensive overview about jitter can be found in an article by J. Schoentgen. The author defines jitter as follows:

> *Jitter designates small, random, involuntary perturbations of the glottal cycle lengths. [Sch01]*

An equivalent definition could be given for the glottal cycle amplitudes. The effect of these variations is similar to the influence of noise upon a signal, except that the variations are an important part of the sound and therefore not unwanted like noise in a signal chain.

1.1.3 Harmonics and noise

Taking Figure 1.2 (cf. page 5) into account, another particular sound property can be seen apart from the harmonics. The spectrum illustrates noise between the harmonics and for frequencies > 4 kHz. Noise dominates in unvoiced phonemes like fricatives but is also present in voiced sounds like vowels. In phoniatrics the ratio of the harmonic and the noisy part (harmonics-to-noise-ratio, HNR) of the signal is an important measure for diagnosis and therapy of voice disorders like a hoarse voice. The origin

of noise in the human voice can be manifold but the physics behind the generation process is similar in all cases and will be described in chapter 4. An accurate description of the noise to harmonic ratio of a voice signal is difficult because of the time-variant fundamental frequency. If the signal was stationary, efficient methods could be applied that are advantageous for the harmonics-noise separation of musical instruments like flue organ pipes [Rio01]. Due to the effect of vibrato and jitter, these methods cannot be used for the singing voice. A discussion of this problem is done in section 4.4.

1.1.4 Choir sound

A choir can be called an arrangement of several singers, some of whom sometimes sing the same notes, and who sing more or less harmonic intervals against other voices.

The particular sound of a choir is often called choral sound and is more than just the addition of separate voices of different or same pitch. S. Ternström has described some of the effects, including those of room acoustics and the feedback control of singers, from a signal processing point of view [Ter91].

A close look at the signal structure of two voices that should follow the same score reveals that the signals are only roughly identical – differences occur in timbre (structure of the harmonics), relative phase and amplitude.

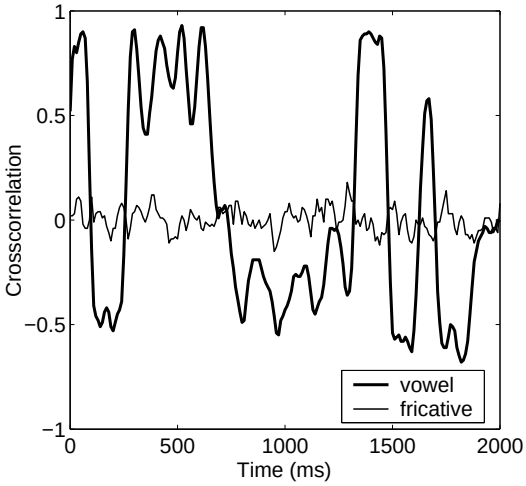


Figure 1.3: Correlation of two voices that sing the sounds [o:] and [ʃ:]

The difference of voiced and unvoiced sounds with respect to the correlation, i. e. the similarity, of the signals is given in Figure 1.3. Two male singers were recorded during simultaneous phonation in the same (anechoic) room with microphones located close to the mouth [JK98]. The crosstalk separation was better than 20 dB. The thick solid line shows the cross-correlation between the sound pressure signals of the two singers, who were asked to sing the sound [o:][3] at same pitch. From the strong fluctuations it can be seen that the correlation alternates between high negative and positive values. This means that the signals are similar but their relative phase alters with time during the adaptation process. The thin solid line indicates the correlation between the pressure signals when

[3] A list of phonemes according to the Association Phonétique Internationale (API) is given in appendix B.

the sound [ʃ:] is pronounced. In this case, the correlation is almost zero, meaning that the signals are not correlated. Music for choirs mostly consists of vowels, i. e. sustained voiced sounds, that are separated by consonants. Therefore, in a choir stationary sounds like vowels will show interference effects between singers in the same group that contribute to the choral sound.

Another important aspect of the choir sound is the radiation characteristics of a singer. This topic will be discussed in chapter 5.

1.2 Singing styles

1.2.1 Registers

There is no general agreement about the classification of the singing voice in registers. Following J. Sundberg [Sun87], the registers will be differentiated according to their homogenity in sound when perceived by a listener. However, the origin for the differentiation of registers is a change in the vocal fold configurations.

Common designations of the voice used for speech and singing with comfortable vocal fold configuration are modal or chest register.

Above the modal register the falsetto register is found in male voices, whereas two different registers can be distinguished in female voices in the upper region: middle and head register. Female voices with very high pitch are often identified as whistle register.

Below the modal register the voice can produce sounds in two additional registers: the straw bass (German: Strohbaß) and the vocal fry register [Hol74]. In Western singing styles these register play a minor role, whereas in Eastern countries the low registers are common. The straw bass register is perceived as a very low but still natural sounding voice, while the vocal fry is different from the above registers because it is based upon a different vocal fold vibration pattern. In the vocal fry or pulse register a periodic interruption of the vocal fold (VF) oscillation cycle can be observed, causing a reduction of the fundamental frequency f_0 down to integer fractions like $f_0/2$, $f_0/3$ etc.

Apart from the registers described above, further terms are used for extreme phonation modes. The vocal fry register differs from a phonation mode being used in the Tibetian chant tradition: the vocal-ventricular phonation mode (VVM) [Fuk99]. L. Fuks claims that in this style the ventricular folds oscillate at an integer fraction of the vocal fold frequency.

1.2.2 Overtone singing

Overtone singing is a technique of forming the overtone spectrum of a sung tone in such a way that one higher harmonic rather than the fundamental frequency determines the melody pitch. Using a sophisticated articulation technique, an overtone singer achieves a perceived separation of the fundamental and one partial. This technique is also called biphonic singing, like the (pathologic) simultaneous generation of two vocal fold modes, most overtone singing styles seem to achieve the separation by modification of the vocal tract only [Hai01]. Although some singers can even produce three separated tones, the following description will be restricted to the generation of one upper and one lower tone. The upper tone is often called melody tone whereas the fundamental is often called drone because it often serves as a musical fundament that changes its low pitch rather slowly. In the following these names will be used for the description of the lower and the upper tone of a biphonic sound. Experienced singers can yield a separation in amplitude between drone and melody tone of more than 40 dB.

Figure 1.4 shows the spectrum of a biphonic sound that has been recorded in an anechoic room at 1 m distance from the singer's mouth[4].

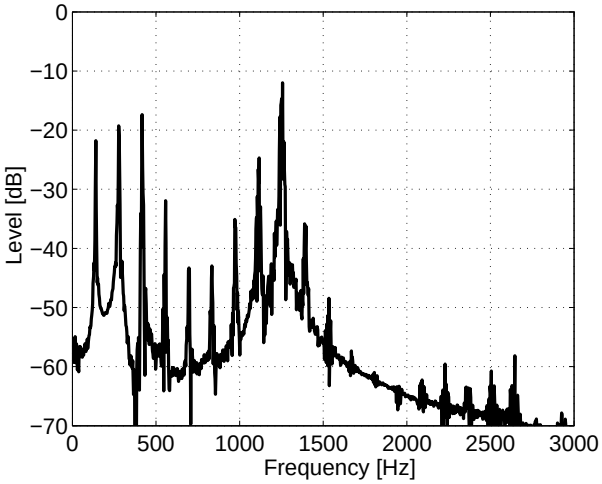


Figure 1.4: Spectrum of a biphonic sound

Only a few scientific publications about overtone singing are available. Extensive studies have been carried out by Trân Quang Hai [Hai00], who describes the broad variety of different overtone styles. A recent study of S. Adachi and M. Yamada [AY99] presents measurements (sound pressure and MRI images) and simulation of Xöömij singing, a special singing style originally used by mongolian singers [HG80]. Adachi supports the "resonance" theory [HG80], which considers the source for the melody tone to be a separated harmonic of the lower tone.

Section 6.3.3 deals with detailed measurements and synthesis of overtone singing.

[4]Sound examples and spectrograms from live recordings can be found at the internet page `http://www.akustik.rwth-aachen.de/~malte/overtone`.

1.3 Discussion

As presented in this chapter, the singing voice is not only a versatile musical instrument but also a generator of a rather complex sound featuring the described properties. From a signal analysis point of view, the voice has been studied intensively in the past. Models for additive synthesis of the singing voice have been built by e. g. P. Cook [Coo90], who describes many details of the voice organ and assembles their effects by connecting sound generators and filters.

The control of parameters for articulatory voice synthesis is another complex topic that has not yet been mentioned. Modern communication devices like GSM mobile phones use sophisticated algorithms to encode and decode speech for bandwidth reduction. These algorithms aim at modeling the most important speech properties by extracting important parameters from the glottal waveform and the acoustic vocal tract properties. Examples are vocoder and code-excited linear prediction (CELP).

In contrast to the approaches above, this thesis aims at describing the synthesis of a voice signal using physical descriptions of the functional components and measured parameters of the voice organ. This method differs significantly from the additive synthesis and speech coding, since the number of parameters necessary for physical models is of similar order as the actual properties of the voice organ. However, voice production using a physical model should give similar results, if the model is sufficiently accurate.

Chapter 2

Vocal folds

The vocal folds (VF) are the most important functional components of the voice organ. Situated in the *larynx*, the VF are functioning as a generator of voiced sounds. Their task is similar to that of a valve modulating the air flow from the lungs to the vocal tract.

2.1 Biomechanics

In Figure 2.1 a schematic graph of the vocal folds is given. The region below the glottis is the so-called subglottal area and consists of the *caudal* surface of the VF and the adjacent trachea that is connected to the broncho-pulmonary organ. Above the VF the supraglottal area is the lower part of the vocal tract (VT), where the ventricular folds are situated. They are often also called false vocal folds, because under certain conditions they can be responsible for voiced sound generation, as explained in section 1.2.1.

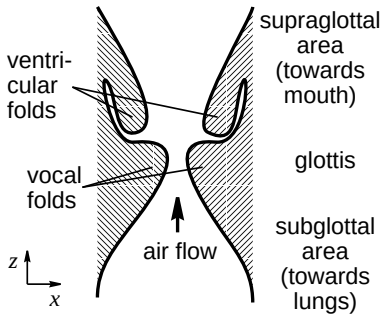


Figure 2.1: Sketch of the glottis, sectional view

Apart from phonation, the VF open the airway for breathing, whereas during phonation the VF are close to one another. These two modes of function are denoted VF abduction (opening of the glottis) for breathing and VF adduction (closing of the glottis) for phonation. The positioning of the VF and therefore the geometric properties of the opening between the vocal folds and the tensions within the folds are determined by the muscles within the vocal folds. The driving force for the oscillation of the VF, however, is not caused by laryngeal muscle action but is a result of the air flow through the constriction of the adducted glottis and the resulting forces on the VF tissue. Historically, the essential functioning of the vocal fold movement has not been clearly understood before 1956 when R. Schilling explained the self-sustained nature of the VF movement

[Sch56]. In 1962 still a theory claimed the oscillation of the VF to be caused by an active control of the *nervus recurrens* [Hus62].

Muscles

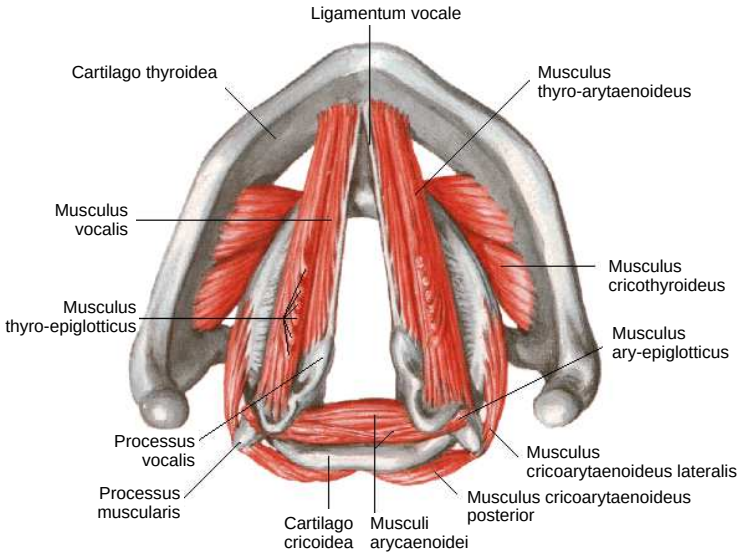


Figure 2.2: Drawing of vocal fold muscles (after [Net99])

In Figure 2.2 the vocal folds and their surrounding muscles and cartilages are shown from above, the top corresponds to the front, the bottom to the rear of the larynx. Three different groups of muscles can be distinguished: muscles that open or close the glottis slits plus muscles stretching the vocal folds. Figure 2.3 depicts the arrangement of the muscles and their effect on the vocal folds. Inside the larynx

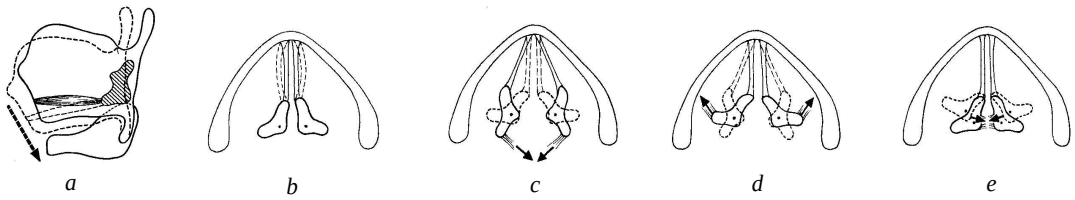


Figure 2.3: Organisation and functions of the muscles in the larynx (after [Boe93])

the vocal folds are stretched between *cartilago thyroidea* and *arytenoid cartilage*. The tension of the vocal folds is adjusted roughly by *musculus cricothyroideus* (2.3a), whereas *musculus vocalis* (2.3b) does the fine-tuning.

Only one muscle can open the glottis slit between the vocal folds, *musculus cricoarytaenoideus posterior* (cf. Figure 2.3c). This muscle is primarily used to allow breathing. The glottis is closed in the front by *musculus cricoarytaenoideus lateralis* (2.3d), whereas the rear part is closed by *musculus arytaenoideus transversus* (2.3e).

Organisation

Figure 2.4 illustrates the sectional view of one VF. The vocal fold consists of the *musculus vocalis*, the *ligamentum vocale* and a mucous membrane epithel, denoted *mucosa*. The tissues are build up in layers, therefore the structure is not homogeneous and not isotropic, either. The slimy *mucosa* is moving on the surface of the *ligamentum vocale*.

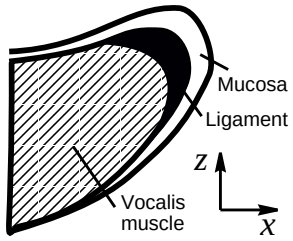


Figure 2.4: Sectional view of one VF (after [Hir68])

Movement

In Figure 2.5 the movement of the VF during one cycle of oscillation is shown, driven by the air flow from below the VF.

The movement is quite complex: the VF move sideways, but there is a an additional action of the upper part that is often called *mucosa*l wave. High-speed recordings or stroboscopic imaging of the vocal folds reveal a phase lag between the lower and the upper part of the VF. This phase lag is essential for the self-sustained oscillation of the VF, because there is a difference in aerodynamics between a convergent (2^{nd} to 4^{th} view from top of Figure 2.5) and a divergent glottis (5^{th} to 7^{th} view).

Technically speaking, the vocal fold movement is driven from the lung pressure and can be understood as a self-sustained exchange of potential and kinetic energy caused by an equilibrium of forces on the vocal fold tissue. These forces can be categorised into two groups, i. e. relatively slowly changing forces that are imposed by quasi stationary boundary conditions like muscle tensions on the one hand and, on the other hand, rapidly changing stress-, damping- or aerodynamic forces that occur during one cycle of oscillation.

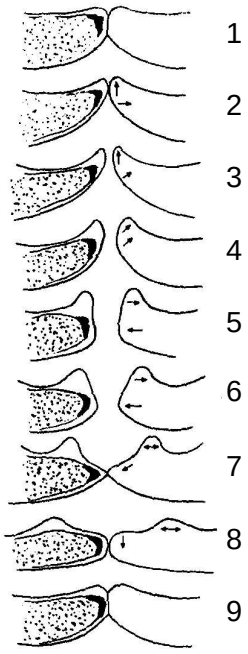


Figure 2.5: Glottal cycle for modal register (taken from [Hir68])

Dimensions within the larynx ($< 2\,\text{cm}$) are small compared to the wavelengths considered for phonation ($> 7\,\text{cm}$). The Strouhal number has been found to be < 1

for speech [Hir92], therefore the flow through the glottis can be assumed to be quasi stationary. In addition, as a first-order approximation, laminar flow and lossless propagation are assumed. Therefore Bernoulli's law can be applied, giving a relation between static pressure P and dynamic pressure $\frac{1}{2}\rho v^2$:

$$P + \frac{1}{2}\rho v^2 = const. \tag{2.1}$$

An application of the Bernoulli equation to the case of a constant air flow through a constriction illustrates the reduction of the pressure P while the particle velocity v is increased. This pressure drop causes forces that pull the VF together.

When the VF are just about to open, the glottis is convergent and no pressure difference across the glottis occurs. Only the quasi-static lung pressure presses the vocal folds apart. When the vocal folds are closing, the *mucosa* is delayed and the flow through the divergent glottis exhibits a profile with slower air movement at the surface, the shear layer, and faster movement at the centre of the flow. As a consequence, flow separation occurs and vortex shedding takes place. These phenomena will be discussed in chapter 4. An excellent review of these and further aeroacoustic phenomena and their application to musical instruments and the human voice has been written by A. Hirschberg in [HKW95].

2.2 Vocal fold models

If the interaction between the vocal folds and the other functional components is considered, a schematic signal flow as given in Figure 2.6 can be drawn.

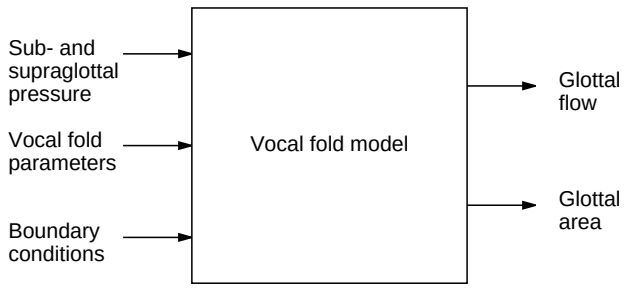


Figure 2.6: Function sketch of a vocal fold model

As input parameters to the VF model, the setting of the glottal muscles and surrounding tissues plus the sub- and supraglottal pressures are important. As output parameters, the glottal area and the glottal flow are generated. The parameters are not independent because the modulation of the air flow significantly changes the sub- and supraglottal pressure conditions. Reflected waves, e. g. from the vocal tract, will change the pressure conditions as well. In section 6.1 these more complex aspects will be considered.

2.2.1 Two-mass models

The use of lumped elements for the characterization of the vocal fold properties is a characteristic aspect of the common two-mass VF model. The main idea of this approach is the reduction of the anisotropic structure of the fold tissue into a small number of condensed elements such as masses, springs and stiffness. In 1969 the first attempt was published of a model that used a one-mass oscillator for each VF [FC69]. The model can produce voiced sounds with additional constraints only [Rod95]. In 1970 the first model with two masses was presented by D. E. Dudgeon [Dud70]. In Figure 2.7 the general set-up of such a two-mass model is shown.

The dashed line indicates the sectional view of the vocal fold. The two masses are denoted m_m and m_v and shall represent the tissues of the VF, not necessarily the *mucosa* and *vocalis* tissues. The coupling of each masses to the bounding cartilages is modelled by a damping element D^b and a spring element k^b. The coupling of both masses to one-another is represented by a spring element k^m without damping. External forces on the masses are described by F_m and F_v. The movement of the masses is often restricted to the $x-$direction and therefore, during movement, only $x-$components of the coupling elements are taken into account.

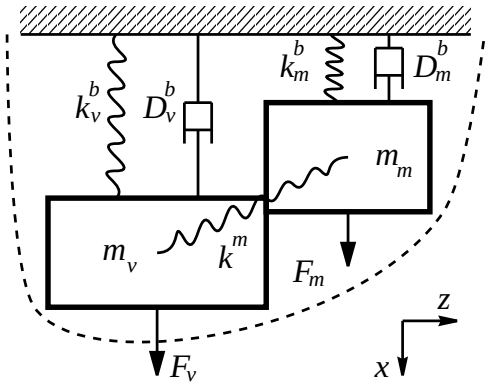


Figure 2.7: Set-up of a two-mass model

A self-oscillating VF model requires at least two degrees of freedom (DOF) because a one-dimensional mass-spring system without acoustic feedback would not be capable of sustained oscillations if constant energy is supplied. However, this does not imply that two DOF are sufficient for a satisfactory description of the vocal fold movement: it is the minimum requirement for a VF model. The movement of the masses can be described by a system of two differential equations:

$$\begin{aligned} m_v\ddot{x}_v + D_v\dot{x}_v + k^b x + k^m(x_v - x_m) &= F_{x,v} \ , \\ m_m\ddot{x}_m + D_m\dot{x}_m + k^b x + k^m(x_m - x_v) &= F_{x,m} \ . \end{aligned} \tag{2.2}$$

In equation (2.2) $m\ddot{x}$ is the inertial force on the mass, $D\dot{x}$ represents the force resulting from the damping of the oscillator, $k^b x$ expresses the spring force that results from the compression or expansion of the spring between mass and boundary, and $k^m(x_v - x_m)$ or $k^m(x_m - x_v)$ is the force that is caused by the coupling of the two masses. The force F_x is the external force driving the oscillator. In the case of the vocal folds, this

aerodynamic force is generated by the static pressure and/or the Bernoulli forces in the glottis. The forces and their origins are discussed exemplarily for the model of Ishizaka and Flanagan in the following section.

Model of Ishizaka and Flanagan

The first model that describes one vocal fold as a system of two coupled masses is the model of K. Ishizaka and J. L. Flanagan [IF72], subsequently called IF model. Their model does not yet assume jet generation at the glottis but is still the most common approach for vocal fold modeling. In Figure 2.8 the set-up of the IF model is shown.

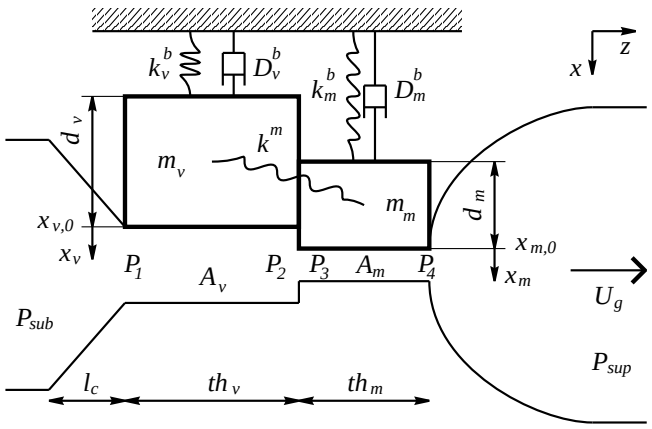


Figure 2.8: Set-up of the IF model

The model is symmetric, as only one vocal fold has been implemented, and the resulting glottal flow U_g is calculated under the assumption of a symmetric VF movement. From the three main sections subglottal, supraglottal and glottal area, the latter is subdivided into four regions where significant pressure changes occur.

Transglottal pressures: The pressure differences across the glottal sections are shown in Figure 2.9.

The first section is characterised by the reduction of the cross-section from the trachea to the area A_v between the masses m_v across the constriction length l_c and yields the following pressure change:

$$P_{sub} - P_1 = 1.37\frac{\rho}{2}\left(\frac{U_g}{A_v}\right)^2 + \int_0^{l_c} \frac{\rho}{A(x)} dx \dot{U}_g. \quad (2.3)$$

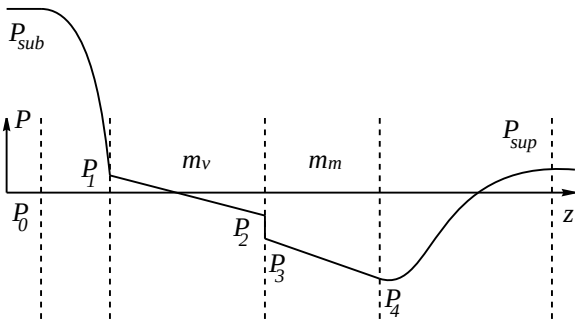


Figure 2.9: Pressure drop across the glottis

The area $A_v = 2l_g(x_{v0} + x_v)$ is calculated from the length of the glottis l_g and the distance x_v of the mass m_v to the neutral position x_{v0}. The area A_m between the *mucosa* masses is calculated accordingly. With $v = \frac{U}{A}$ the first term in (2.3) is

identical to the Bernoulli law (2.1) with exception of a leading factor. This factor is based upon the assumption of the *vena contracta* effect, which describes the increase of the pressure drop at sharp discontinuities. Measurements of van den Berg [vdBZD57] on gypsum models of the larynx determined the factor to be 1.37. However, the assumption of a sharp edge at the entry of the glottis has been discussed controversely [P$^+$94]. The second term in (2.3) takes into account the geometry change in the constriction.

In the following glottal section the pressure decreases linearly over the mass m_v due to viscous losses that are characterised by the shear viscosity coefficient μ. The pressure difference across the thickness th_v of the *vocalis* mass in $z-$direction is then

$$P_1 - P_2 = 12\frac{\mu a^2 th_v}{A_v^3}U_g + \frac{\rho th_v}{A_v}\dot{U}_g \ . \tag{2.4}$$

Because of the flow continuity across the discontinuity, the pressure jumps according to (2.1) at the transition between both masses.

$$P_2 - P_3 = \frac{\rho}{2}U_g^2\left(\frac{1}{A_m^2} - \frac{1}{A_v^2}\right). \tag{2.5}$$

Across the mass m_m again a linear pressure decrease as in (2.4) takes place.

$$P_3 - P_4 = 12\frac{\mu a^2 th_m}{A_m^3}U_g + \frac{\rho th_m}{A_m}\dot{U}_g \ . \tag{2.6}$$

In the last glottal section the pressure "recovers" towards the ambient atmospheric pressure outside the voice organ. This pressure difference is

$$P_4 - P_{sup} = -\frac{\rho}{2}\left(\frac{U_g}{A_m}\right)^2 k_e \ , \tag{2.7}$$

with the pressure recovery coefficient k_e

$$k_e = 2\frac{A_{min}}{A_{sup}}\left(1 - \frac{A_{min}}{A_{sup}}\right). \tag{2.8}$$

The factor k_e shall describe the expansion of the flow to the vocal tract walls. However, the diameters of the glottal area and the vocal tract entry area are quite different during phonation ($A_{min} \ll A_{sup}$), thus the factor is often neglected in recent models [PHWB95].

Aerodynamic forces: The aerodynamic forces F^a on the masses m_v and m_m are calculated as averaged pressures on the areas of the masses. The areas are calculated from the *vocalis* thickness d_v or the *mucosa* thickness d_m and the length l_g. Three cases need to be distinguished:

1. In case of an open glottis the following equations apply:
$$\begin{aligned} F_v^a &= \tfrac{1}{2}(P_1 + P_2) d_v l_g \ , \\ F_m^a &= \tfrac{1}{2}(P_3 + P_4) d_m l_g \ . \end{aligned} \tag{2.9}$$

2. If only the masses m_m are closed, the dynamic pressure disappears and only the static pressure is present: $P_1 = P_2 = P_{sub}$. As a consequence the forces are
$$\begin{aligned} F_v^a &= P_{sub} d_v l_g \ , \\ F_m^a &= 0 \ . \end{aligned} \tag{2.10}$$

3. In the remaining cases, when only the masses m_v or all masses are closed, no aerodynamic forces act on the masses:
$$F_v^a = F_m^a = 0 \ . \tag{2.11}$$

Spring forces: The forces that act on the springs between the masses and the boundary are different for the cases of the open and the closed glottis. For the open glottis the force is given to
$$F^b = k_o \left(x + \eta_{k_o} x^3 \right) . \tag{2.12}$$
k_o is the linear spring stiffness for the open vocal folds. The coefficients η_{k_o} and η_{k_c} describe the nonlinear character of the springs. With the linear spring stiffness k_c of the closed vocal folds the following equation applies to the case of the closed glottis:
$$F^b = k_o \left(x + \eta_{k_o} x^3 \right) + k_c \left[(x + x_0) + \eta_{k_c} (x + x_0)^3 \right] . \tag{2.13}$$
The coupling of the masses one to the other yields the following force on the masses:
$$\begin{aligned} F_v^m &= k^m (x_v - x_m) \ , \\ F_m^m &= k^m (x_m - x_v) \ . \end{aligned} \tag{2.14}$$

Damping forces: The damping increases when the glottis is closed (see eq. (4) in [IF72]). The damping coefficient is then
$$D = 2\xi \sqrt{mk} \ . \tag{2.15}$$
In this equation, ξ denotes the degree of damping of an uncoupled oscillator. For $\xi < 1$, the system oscillates, whereas for $\xi > 1$ the system is damped too much (above critical damping) (cf. [Vog95]).

Mass movement: The movement of the masses is described by a system of differential equations.
$$\begin{aligned} m_v \ddot{x}_v + D_v \dot{x}_v + F_v^b + F_v^m &= F_v \ , \\ m_m \ddot{x}_m + D_m \dot{x}_m + F_m^b + F_m^m &= F_m \ . \end{aligned} \tag{2.16}$$
The equations are iteratively solved using start values of the rest positions.

Smooth models

The idea of a smooth border line between the "edges" of the VF has been described by D. G. Childers, who introduces a one-mass model that has been applied to the simulation of voice pathologies and synthesis of vocal fry [CHMA86].

The characteristics of a free jet that separates from the VF surface, and the evidence of a moving separation point have been described in theory and measurements by X. Pelorson *et al.* [P+94]. As no discontinuities in the pressure distribution over the glottis occur, the forces on the VF outline can be calculated analytically.

These aspects have been implemented into a two-mass model by N. Lous *et al.* [LHVH98]. Continuous lines between the boundaries and both point masses are assumed as shown in Figure 2.10.

This is in contrast to the IF model that uses stepwise geometry changes.

Another new aspect of Lous' model is that, instead of a combination of a small and a big mass, two identical oscillators on each side of the glottis are used. Therefore, a reduction of parameters can be achieved: $m_1 = m_2$, $k_1^b = k_2^b$, $D_1^b = D_2^b$.

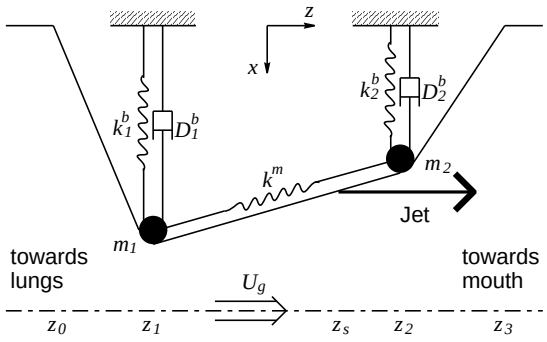


Figure 2.10: Geometry of Lous' model

The time-dependent channel height h in the sections between the co-ordinates of the smooth model $i = 0..3$ is calculated as follows:

$$h(z) = \frac{h_i - h_{i-1}}{z_i - z_{i-1}}(z - z_{i-1}) + h_{i-1} \ . \tag{2.17}$$

Lous assumes a quasi-stationary frictionless and incompressible flow between the trachea and the point z_s, where the jet separates from the connection line between the masses [LHVH98]. The pressure difference between the subglottal pressure P_{sub} and the pressure in the glottis $P(z)$ at position z is calculated using Bernoulli's equation:

$$P_{sub} - P(z) = \frac{\rho_0}{2l_g^2}\left(\frac{1}{h(z)} - \frac{1}{h_0}\right)U_g^2 \ . \tag{2.18}$$

C. Vilain *et al.* [VPT99, Vil01, VPH+01] describe the flow through symmetrical and asymmetrical replicas of the VF and focus on a description of the glottal flow in both cases. In contrast to Lous' model the two oscillators for each VF are not necessarily identical, and the model takes into account viscous losses (Poiseuille flow) and the

instationarity of the flow by adding two terms to the pressure equation (2.18):

$$\begin{aligned} P_{sub} - P(z) &= \tfrac{\rho_0}{2l_g^2}\left(\tfrac{1}{h(z)^2} - \tfrac{1}{h_0^2}\right) U_g^2 \\ &+ \tfrac{12\mu}{2l_g}\left[\tfrac{z_1-z_0}{h_1-h_0}\left(\tfrac{1}{h_0^2} - \tfrac{1}{h_1^2}\right) + \tfrac{z_2-z_1}{h_2-h_1}\left(\tfrac{1}{h_1^2} - \tfrac{1}{h_s^2}\right)\right] U_g \\ &+ \tfrac{\rho_0}{l_g}\left[\tfrac{z_1-z_0}{h_1-h_0}\ln\left|\tfrac{h_1}{h_0}\right| + \tfrac{z_2-z_1}{h_2-h_1}\ln\left|\tfrac{h_s}{h_1}\right|\right] \dot{U}_g \ . \end{aligned} \tag{2.19}$$

For both models, the forces acting on each of both masses $m_i, i = 1, 2$ are calculated as follows:

$$F_i^a(t) = \int_{z_{i-1}}^{z_i} \left(\frac{z - z_{i-1}}{z_i - z_{i-1}}\right) P(z)\mathrm{d}z + \int_{z_i}^{z_{i+1}} \left(\frac{z_{i+1} - z}{z_{i+1} - z_i}\right) P(z)\mathrm{d}z \ . \tag{2.20}$$

Modified two-mass models

Modifications of the IF model have been done by T. Koizumi *et al.* [KTH87] with respect to the geometric properties of the masses and the coupling spring properties. The models aim at a more realistic voice synthesis. However, the pressure drop across the glottis is not significantly different from that of the original IF model.

The model of I. Steinecke and H. Herzel [SH95] has been developed for studies of voice disorders using methods from nonlinear dynamics. The IF model has been reduced using the findings discussed in [P+94] to a basic description of the resonators, leaving just the coupling, linear springs and the Bernoulli forces. Due to its simplicity, the model has often been used for parameter studies (cf. e. g. [TMH+97, JZ01]), and has recently been extended to a three-dimensional surface [Dre01].

With calculations based upon a simplified IF model, J. C. Lucero [Luc96] showed that register changes between chest and falsetto register can be associated with a transition of the spring stiffness k_m^b across a critical value. This interpretation does not necessarily rely on physical reasoning but indicates the flexibility of the IF model.

2.2.2 Other models

Few other models have been described that differ significantly from the lumped-element approaches based upon the IF model. However, some approaches are briefly described in the following.

Multiple mass models

In 1973 I. R. Titze extended the IF model by dividing each mass in a number of 8 masses in longitudinal ($y-$) direction ($l_g = 8a$) [Tit73, Tit74], subsequently called T73 model.

A schematic drawing of the set-up is shown in Figure 2.11. The main advantage of such an approach is the possibility to model physical modifications and waves in the longitudinal dimension of the vocal folds. The model is described in detail in section 2.3.

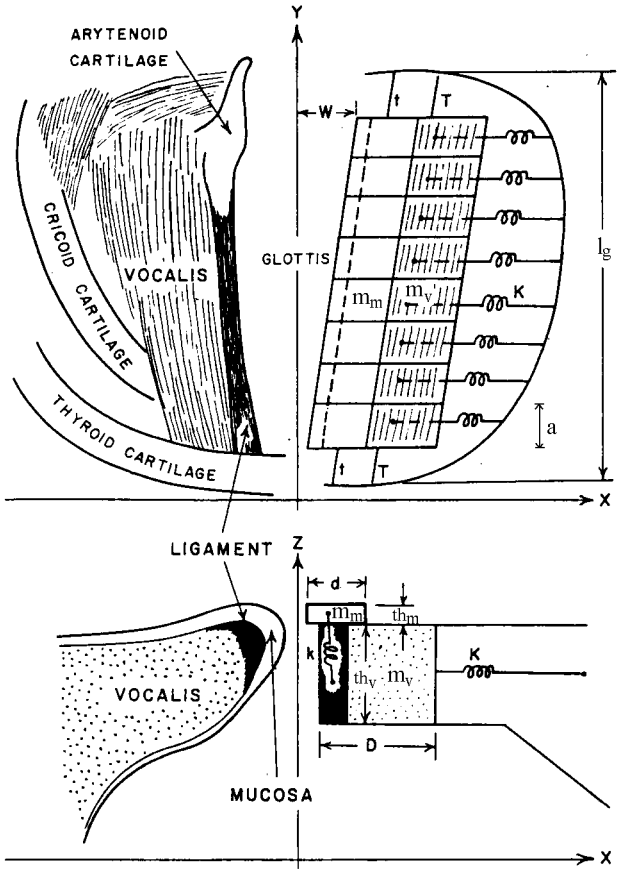


Figure 2.11: Schematic set-up of the 16-mass model (after [Tit73])

The model of D. Wong *et al.* [WICT91] is a hybrid model that combines the IF72 and the T73 model because it divides the two masses in each five segments in longitudinal direction. However, the model does not include jet generation.

In 1995 B. Story and I. R. Titze presented the body-cover model [ST95], that adds another mass to the IF model. One big mass is coupled to the boundary (*thyroid cartilage*) and to each of the other two small masses by non-linear springs and damping elements. The small masses are coupled with a linear spring. In this model, the generation of a free jet is assumed. The jet separates from the minimum glottal area. Another difference to the IF model is the neglection of viscous losses and the pressure drop due to the inertance of the air.

Rotational model

Instead of using two masses with one translational DOF each, the rotating mass model of J. Liljencrants [Lil91] replaces one translating mass by a rotational DOF of the one mass. Another new idea is the description of the mass surfaces as rounded areas. The model assumes a free jet that separates from the surface. Depending on the jet diameter, a pressure recovery coefficient is assumed that varies from 0.21 to 0.69. The model aims at a simple description of the complex VF movement rather than at a most physical modeling of all parameters involved.

Mechanical models

A physical VF model can also be a replica of the biomechanic setup of the vocal folds. Most models are optimised for the examination of basic physical aspects like transient flow characteristics [HPH+96] or fold tissue investigations [AT91]. H. Saweda and S. Hashimoto [SH00] built a completely mechanical vocal fold model. Only few mechanical models are suitable for the synthesis of the human voice.

Fluid dynamical models

Numerical approaches to solve the Navier-Stokes equations for the glottal flow have been demonstrated by e. g. J. Liljencrants [Lil89] and X. Pelorson *et al.* [PLK95], who calculated the particle movement through a 2-dimensional channel with a glottis-like geometry. However, a 3-dimensional solution of the Navier-Stokes equations is not possible, and the calculation with dynamically changing boundary conditions is numerically very expensive.

Another approach for the description of the fluid dynamic characteristics of the flow through a constriction like the glottis is the vortex-blob method [Hof98]. The method requires a two-dimensional, incompressible and inviscid flow and high Reynold numbers. The Reynold number is an index for the significance of viscous forces in the flow [HPH+96]. However, it enables the visualization of the vortex shedding and the flow separation at the glottis. Therefore, it is mainly used for investigation of basic flow behaviour rather than for voice synthesis.

Finite element models

In 1996 F. Alipour and I. R. Titze presented a two-dimensional model for the airflow and vocal fold movement [AT96].

Recently, F. Alipour and D. A. Berry presented a finite element (FE) model of the vocal folds [AB01]. A finite element model is very different from the 2-mass approach because it does not assume lumped elements but describes geometry and physical properties of the VF tissue with a huge number of very small elements. With respect to accuracy of the VF description, a FE model is a very good solution if all properties of all elements and their relations are known. Two major drawbacks are associated with such models: due to their complexity, the calculation times for dynamic modeling of VF motion is quite high. Therefore a real-time calculation of the movement is not possible, yet. The other problem is the availability of the huge number of parameters for the accurate description of the vocal folds. As described in section 2.1, the VF consists of several layers of tissue that prove the inhomogeneity and anisotropy of the VF structure. A correct modeling would require knowledge of stress-strain relations in all dimensions of the VF.

2.3 Implemented model

The model presented in this thesis is based on the 16-mass model developed by I. R. Titze [Tit73] but includes some more recently published modifications. These additions are marked with **bold text** and will be summed up in section 2.3.5. The values of parameters used in the equations are discussed in section 2.3.4.

The implemented model differs from the models that are described in the previous section in the following features:

- Modeling of modal, head and falsetto register is possible
- Variation of individual masses or spring stiffness in an arbitrary number of sections allows simulation of e. g. singer's nodules
- All parameters can be changed and pressure or flow values can be monitored during calculation

Set-up: Figure 2.12 illustrates the geometry of the model.

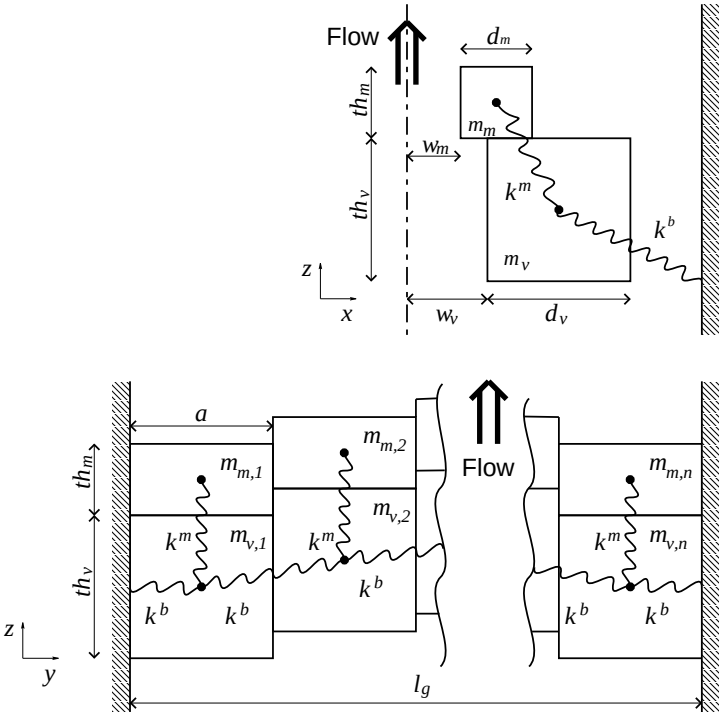


Figure 2.12: Sectional view (top) and side view (bottom) of the implemented model

The arrangement of n masses $m_{m,i}$, with $i = 1..n$ represents the *vocalis* muscle, $m_{v,i}$ represents the mucosa membrane. Each of the big masses $m_{v,i}$ is connected to

the boundary by a spring with stiffness k^b and damping D^b, and to the small mass $m_{m,i}$ by a spring with stiffness k^m **and damping** D^m.

The most important difference between this model and the IF model is the segmentation of the VF in n segments. The sum of the **arbitrary number of** n **segments** of width a is the longitudinal length l_g of the VF. There are no springs between the border and the small masses m_m, which is more close to the actual configuration of the vocal folds, because the mucosa membrane is not directly attached to the *cricoid cartilage*. In Titze's model, the number of 16 masses is fixed.

2.3.1 Forces

In direction of the co-ordinate x, six different kinds of forces F act on each *vocalis* mass $m_{v,i}$. One more force, the gravitation force $m_v \cdot g$ with $g = 9.81\,m\,s^{-2}$ is neglected as in [Tit73]. This can be justified by the fact that the effect of the gravitation force can be understood as an offset of the rest positions of the masses.

Four of the six forces on each *vocalis* mass $m_{v,i}$ are the following:

1. forces due to tensions between a mass and the two neighboured *vocalis* masses $m_{v,i-1}$ and $m_{v,i+1}$ (denoted by a superscript index T),
2. spring forces to the coupled *mucosa* masses $m_{m,i}$ (index m),
3. spring forces to the bordering *cricoid cartilage* (index b), and
4. contact forces during closure of the glottis (index c).

The forces on the mucosa masses are described in appendix D because of better readability of the text.

Tension forces

Titze uses an exponential relation between tension T_l and the resulting strain (elongation) $S_l = (r_l - r_{l,0})/r_0$ of the *ligamentum vocale* tissue that was measured by van den Berg [vdB60]:

$$S_l = S_{l,max}\left(1 - e^{\frac{-T_l}{\tau_l}}\right), \tag{2.21}$$

with $S_{l,max} = 0.3$ as the maximum strain and the coefficient τ_l approximates 350 kPa. Similar equations with exponential stress-strain relations were derived for the *vocalis* and *mucosa* tissues. If the equation is solved for T_l and multiplied with the surface $th_l d_l$ of the ligament section, F_l^T is the tension force on the ligament:

$$F_l^T = -th_l d_l \tau_l \ln\left(1 - \frac{S_l}{S_{l,max}}\right). \tag{2.22}$$

The nonlinear relations beween stress and strain in the *vocalis* and *mucosa* tissues have been determined more detailed by F. Alipour-Haghighi and I. R. Titze [AT91] by measurements on excised larynces of dogs for stress-strain curves for $0\% < S < 40\%$. The VF of dogs do not have vocal cords (*ligamenta vocales*) but are very similar to human VF. The **nonlinear stress-strain curve** is expressed by the **polynomial** equation (2.23).

$$T_v^T = \left(0.4 + 42.3S_v - 341.9S_v^2 + 1132S_v^3\right) \text{kPa}. \tag{2.23}$$

The force $F_{v,i}^T$ is a sum of the forces F_l^T on the *ligamentum vocale* and the *vocalis* stress T_v^T and active stress $T_{v,act}$ that act on the surfaces $th_v d_v$ of the *vocalis*:

$$F_{v,i}^T = F_l^T + th_v d_v (T_v^T + T_{v,act}) \ . \tag{2.24}$$

The active stress $T_{v,act}$ is caused by the action of the muscles *cricothyroideus* and *vocalis* and is one of the most important parameters for the control of the model. The tension forces in $x-$direction between the neighboured *vocalis* masses is then given to:

$$F_{x,v,i}^T = F_{v,i}^T \frac{x_{v,i-1} - x_{v,i}}{r_{v,i}} + F_{v,i+1}^T \frac{x_{v,i+1} - x_{v,i}}{r_{v,i+1}} \ , \tag{2.25}$$

where r denotes the distance to the neighboured element. In $z-$direction, a similar equation applies:

$$F_{z,v,i}^T = F_{v,i}^T \frac{z_{v,i-1} - z_{v,i}}{r_{v,i}} + F_{v,i+1}^T \frac{z_{v,i+1} - z_{v,i}}{r_{v,i+1}} \ . \tag{2.26}$$

Spring forces

The spring forces between the *vocalis* and the *mucosa* mass that are applied in the implemented model differ from the 16-mass model in the **nonlinearity of the connecting springs** and the **differentiation of the spring stiffnesses in $x-$ and $z-$direction**.

$$F_{x,v,i}^m = -k_x^m \left((x_{v,i} - x_{m,i}) + \eta_{k,x}^m (x_{v,i} - x_{m,i})^3\right) \ . \tag{2.27}$$

The nonlinearity factor η takes into account the saturation effect of the spring for large elongations. A similar equation applies for the force in $z-$direction, with exception of a shifted rest position $z_{v,m,0}$ of the *mucosa* mass relative to the *vocalis mass*:

$$F_{z,v,i}^m = -k_z^m \left((z_{v,i} - z_{m,i} + z_{v,m,0}) + \eta_{k,z}^m (z_{v,i} - z_{m,i} + z_{v,m,0})^3\right) \ . \tag{2.28}$$

The forces in $x-$ and $z-$direction to the bounding *cartilago cricothyroideus* are nonlinear as well:

$$F_{x,v,i}^b = -k_x^b \left((x_{v,i} - x_{0,v,i}) + \eta_{k,x}^b (x_{v,i} - x_{0,v,i})^3\right) \ . \tag{2.29}$$

$$F_{z,v,i}^b = -k_z^b \left((z_{v,i} - z_{0,v,i}) + \eta_{k,z}^b (z_{v,i} - z_{0,v,i})^3\right) \ . \tag{2.30}$$

Contact forces

The contact forces represent the non-linear change of stiffness of the vocal fold tissue during the closed phase. In $x-$direction

$$F^c_{x,v,i} = \begin{cases} k^c_{x,v}\left(w_v + \eta^c_{k,x,v} w_v^3\right) & \text{for } w_v < 0\ ; \\ 0 & \text{else.} \end{cases} \tag{2.31}$$

η denotes the nonlinearity coefficient for the stress-strain curves. A similar equation applies for the contact force in $z-$direction:

$$F^c_{z,v,i} = \begin{cases} k^c_{z,v}\big(z_{m,i} - z_{v,i} - z_{v,m,0} \\ + \eta^c_{k,z,v}\left(z_{m,i} - z_{v,i} - z_{v,m,0}\right)^3\big) & \text{for } z_{m,i} < z_{v,i} + \frac{th_m + th_v}{2}\ ; \\ 0 & \text{else.} \end{cases} \tag{2.32}$$

Damping forces

The fifth force is given by the damping force

$$F^d_{x,v,i} = -D^T_{v,i}\dot{d}(x_{v,i} - x_{v,i-1}) - D^T_{v,i+1}\dot{d}(x_{v,i} - x_{v,i+1}) - D^m_{x,i}\dot{d}(x_{v,i} - x_{m,i}) - D^b_{x,i}\dot{x}_{v,i} \tag{2.33}$$

in $x-$direction, and

$$F^d_{z,v,i} = -D^T_{v,i}\dot{d}(z_{v,i} - z_{v,i-1}) - D^T_{v,i+1}\dot{d}(z_{v,i} - z_{v,i+1}) - D^m_{z,i}\dot{d}(z_{v,i} - z_{m,i}) - D^b_{z,i}\dot{z}_{v,i}\ . \tag{2.34}$$

in $z-$direction. The damping coefficients for the coupling springs to the neighboured elements are

$$D^T_{v,i} = 2\xi_{v,i}\sqrt{m_v \frac{F^T_{v,i}}{r_{v,i}}}\ . \tag{2.35}$$

The damping coefficients for the connection to the mucosa mass in $x-$ and $y-$direction are

$$D^m_{x,i} = 2\xi^m_{v,i}\sqrt{m_m k^m_{x,i}} \text{ and } D^m_{z,i} = 2\xi^m_{v,i}\sqrt{m_m k^m_{z,i}}\ , \tag{2.36}$$

and to the border

$$D^b_{x,i} = 2\xi^b_{x,i}\sqrt{m_v k^b_{x,i}} \text{ and } D^b_{z,i} = 2\xi^b_{z,i}\sqrt{m_v k^b_{z,i}}\ . \tag{2.37}$$

This approach differs from Titze's model, where all springs are represented as one equivalent spring constant that sums up the effect of all springs:

$$k_{x,v,i} = \frac{F_{T,v,i}}{r_{v,i}} + \frac{F_{T,v,i+1}}{r_{v,i+1}} + k^m + k^b. \tag{2.38}$$

Aerodynamic forces

The last force is the aerodynamic force F^a. Its value depends on the pressure upon the mass and the absolute and relative position of the masses. In direction of the co-ordinate z, the following equation applies for all cases:

$$F^a_{z,v,i} = P_{sub} d_v a \ . \tag{2.39}$$

The following cases must be distinguished for the *vocalis* mass in direction of the co-ordinate x:

$$F^a_{x,v,i} = \begin{cases} P_v t h_v a & \text{for } A_m < A_v \text{ (convergent)}, \quad \text{open glottis;} \\ P_v t h_v a & \text{for } A_m > A_v \text{ (divergent)}, \quad \text{open glottis;} \\ P_{sub} t h_v a & \text{for } A_m = 0, A_v > 0, \quad \text{closed glottis;} \\ 0 & \text{for } A_m > 0, A_v = 0, \quad \text{closed glottis;} \\ 0 & \text{for } A_m = A_v = 0, \quad \text{closed glottis.} \end{cases} \tag{2.40}$$

The pressures P_v and P_m on the vocalis mass and the *mucosa* mass are derived in the following section.

2.3.2 Pressures

In this implementation **jet separation** is assumed, and therefore **only two different glottal sections** are distinguished. This is in contrast to Titze's 16-mass model, where the pressure distribution of [IF72] (cf. page 18) is adapted with modifications. In a more recent paper [ST95], B. Story and I.R. Titze describe the pressure distribution that applies when jet generation is assumed. The following assumptions are made in [ST95]:

- Jet separation takes place at the beginning of the masses that determine the minimum area of the glottis.
- For the convergent case the Bernoulli equation is applied to the first section (subglottal to minimum area).
- The jet diameter and the pressure in the jet remain constant within the glottis.
- Pressure recovery takes place according to [IM72].

In [ST95] the pressure P_v on the *vocalis* mass are calculated under the assumption of flow separation at the smallest area A_{min}. In the divergent case the whole glottis

is within the jet regime whereas in the convergent case only the *mucosa* mass resides in the jet regime. For the pressures the following equations apply:

$$P_v = \begin{cases} P_{sub} - (P_{sub} - P_{sup}) \frac{\frac{1}{A_v^2} - \frac{1}{A_{sub}^2}}{\frac{1-k_e}{A_{min}^2} - \frac{1}{A_{sub}^2}} & \text{convergent glottis;} \\ P_{sup} - (P_{sub} - P_{sup}) \frac{\frac{k_e}{A_{min}^2}}{\frac{1-k_e}{A_{min}^2} - \frac{1}{A_{sub}^2}} & \text{divergent glottis.} \end{cases} \tag{2.41}$$

The pressure on the *mucosa* mass is

$$P_m = P_{sup} - (P_{sub} - P_{sup}) \frac{\frac{k_e}{A_{min}^2}}{\frac{1-k_e}{A_{min}^2} - \frac{1}{A_{sub}^2}} \ . \tag{2.42}$$

Note that the same pressure applies for the cases of convergent and divergent glottis.

An illustration of the variables in (2.41) and (2.42) is given in Figure 2.13. P_{sub} and P_{sup} are the sub- and supraglottal pressures, A_{sub} and A_{sup} are the sub- and supraglottal areas, A_{min} is the minimum area in the glottis, k_e is the pressure recovery coefficient as defined in (2.8), and A_v is the area between the *vocalis* masses.

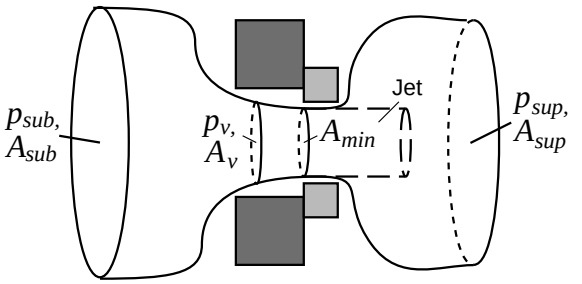


Figure 2.13: Geometry and pressures at the glottis

From these assumptions the following modifications are made for the present implementation:

- The **jet separation point is fixed** at the discontinuity between the masses. Looking at Figure 2.5 most of the *mucosa* mass is located downstream of the minimum glottal area for all instants of the glottal cycle.

- The jet separates the dynamic pressure from the *mucosa* surface. As a consequence, **no Bernoulli forces** but **only the stationary supraglottal pressure act on the *mucosa* masses**.

The assumptions above yield the following equations for all glottal geometries:

$$\begin{aligned} P_v &= P_{sub} - (P_{sub} - P_{sup}) \frac{\frac{1}{A_v^2} - \frac{1}{A_{sub}^2}}{\frac{1-k_e}{A_{min}^2} - \frac{1}{A_{sub}^2}} \ , \\ P_m &= P_{sup} \ . \end{aligned} \tag{2.43}$$

2.3.3 Glottal flow

The volume flow through the glottis is calculated from the pressure drop across the glottis opening. The flow is relevant for the acoustic pressure modulation at the glottis exit.

Figure 2.14 shows the setup of pressures near the glottis when the VF model is inserted into a waveguide.

The index '+' denotes pressure acoustic waves p travelling towards the mouth whereas variables with index '−' indicate the opposite direction. The resulting pressure below the glottis P_{sub} is the sum of the waves to and from trachea $p^{\pm}_{sub}$ and the nearly stationary lung pressure P_{lung}

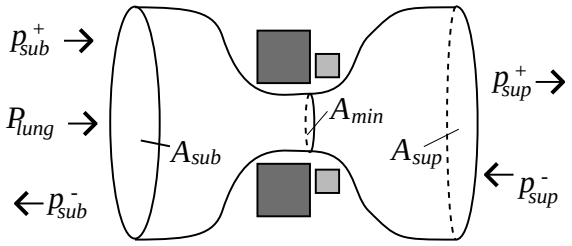


Figure 2.14: Geometry and pressures in the waveguide

$$P_{sub} = p^+_{sub} + p^-_{sub} + P_{lung} \ . \quad (2.44)$$

The pressure P_{sup} above the glottis is

$$P_{sup} = p^-_{sup} + p^+_{sup} \ . \quad (2.45)$$

In the glottal section between subglottal and minimum glottal area, the pressure is determined by the Bernoulli equation (2.1). At A_{min} the following equation applies for the pressure:

$$P_{A_{min},Bern.} = P_{sub} - \frac{1}{2}\rho U_g^2 \left(\frac{1}{A^2_{min}} - \frac{1}{A^2_{sub}} \right) . \quad (2.46)$$

Downstream, the acoustic pressure is calculated differently from the stationary pressure in (2.43):

$$P_{A_{min},jet} = P_{sup} - \frac{1}{2}\rho k_e U_g^2 \frac{1}{A^2_{min}} \ . \quad (2.47)$$

The combination of (2.46) and (2.47) yields the transglottal pressure P_{tg}:

$$P_{tg} = P_{sub} - P_{sup} = \frac{1}{2}\rho k_e U_g^2 \left(\frac{1 - k_e}{A^2_{min}} - \frac{1}{A^2_{sub}} \right) . \quad (2.48)$$

Flow continuity enforces the same flow up- and downstream from the glottis:

$$\begin{aligned} U_g &= \tfrac{A_{sub}}{\rho c}(p^+_{sub} - p^-_{sub}) \text{ and} \\ U_g &= \tfrac{A_{sup}}{\rho c}(p^+_{sup} - p^-_{sup}) \ . \end{aligned} \quad (2.49)$$

With the abbreviations $k_t = 1 - k_e - (A_{min}/A_{sub})^2$ and $\frac{1}{A^*} = \frac{1}{A_{sub}} + \frac{1}{A_{sup}}$, the flow in the glottis U_g can be written as a combination of the equations (2.49), (2.44) and (2.48):

$$U_g = \frac{cA_{min}}{k_t} \left\{ -\frac{A_{min}}{A^*} \pm \sqrt{+\left(\frac{A_{min}}{A^*}\right)^2 + 2\frac{k_t}{c^2\rho}\left(2p^+_{sub} + P_{lung} - 2p^-_{sup}\right)} \right\} . \quad (2.50)$$

The acoustic pressure waves that leave the glottis in direction of the lungs and the vocal tract can be calculated with (2.50) and (2.44/2.45):

$$\begin{aligned} p^-_{sub} &= p^+_{sub} + \tfrac{\rho c}{A_{sub}} U_g \ , \\ p^+_{sup} &= p^-_{sup} + \tfrac{\rho c}{A_{sup}} U_g \ . \end{aligned} \quad (2.51)$$

It is assumed that the flow-induced pressure waves $\frac{\rho c}{A_{sup}} U_g$ generate an acoustic point source. The produced sound wave p^+_{sup} does not contribute to the calculations within the glottis since the jet separates it from the static pressure downstream of the glottis. Upstream of the glottis, no reflected waves from the lungs have been taken into account, yet.

Mass movement: The motion of the masses can be described according to Newton's law:

$$m_v \ddot{x}_{v,i} = F^T_{x,v,i} + F^m_{x,v,i} + F^b_{x,v,i} + F^c_{x,v,i} + F^d_{x,v,i} + F^a_{x,v,i} \ . \quad (2.52)$$

For the *mucosa* mass, the equations are given in appendix D. A detailed description of the implementation of the model can be found in the thesis by Nils Alhäuser [Alh99].

2.3.4 Parameters

Physical modeling of real structures is limited by the accuracy of the parameters used as well as by simplification of physical laws. For the vocal folds the parameters are difficult to obtain because direct, *in vivo*, measurements of material properties such as tissue density and stiffness are impossible. Some properties of the vocal folds such as tensions or geometric dimensions can be measured with accuracy, but others are difficult to obtain, such as values for damping [HMOI92, IK68]. Most parameters have been derived indirectly [IK68] by measurements on the human voice organ and some have been derived from excised canine larynces.

Masses

The masses are calculated from the dimensions and the tissue densities derived from measurements of excised larynces.

$$\begin{aligned} m_{v,total} &= \underbrace{th_{voc.} d_{voc.} l_g \rho_v}_{M_{voc.,total}} + \underbrace{th_l d_l l_g \rho_l}_{m_{lig.,total}} , \\ m_{m,total} &= th_m d_m l_g \rho_m . \end{aligned} \tag{2.53}$$

For non-pathologic configurations the total mass is distributed homogeneously over the vocal fold tissues. Therefore, each of the n_m mass segments has the fractional mass

$$\begin{aligned} m_v &= \tfrac{m_{v,total}}{n_m} , \\ m_m &= \tfrac{m_{m,total}}{n_m} . \end{aligned} \tag{2.54}$$

In Table 2.1 the geometrical and physical values for the fold parameters are listed. The geometrical and mass properties differ significantly for different registers of phonation. The modifications from the modal registers are explained in detail in section 6.3.1.

Table 2.1: Geometrical and physical properties of the vocal fold model, configuration for modal register

	Length [mm]	Width [mm]	Thickness [mm]	Density [mg/mm^3]	Mass [mg]
Vocalis	$l_g = 14$	$d_v = 2.5$	$th_v = 2.5$	$\rho_v = 1.04$	$m_{voc.,total} = 91$
Ligamentum vocale	$l_g = 14$	$d_l = 1.0$	$th_l = 1.0$	$\rho_l = 1.04$	$m_{lig.,total} = 15$
Mucosa	$l_g = 14$	$d_m = 1.0$	$th_m = 0.5$	$\rho_m = 1.02$	$m_{m,total} = 7$

Glottal gap

In the rest position of the VF, the glottal gap is responsible for the initial VF movement, a possible glottal leakage, and the open quotient (OQ) of the oscillation cycle. If the VF are closed in the rest position, no oscillation would start because no air flow exists. If the VF are too distant from each other, the VF would not start a regular cyclic movement because the VF would never close. The open quotient is the ratio between the amount of time in each cycle when the glottis is open and the total oscillation period. With the help of high-speed recordings of the human VF, the OQ has been measured to be 57..84 % for healthy subjects [CHM$^+$90]. The OQ differs with the singing style and varies significantly in pathologic voices. With the standard configuration of the developed model[1], a glottal gap of

$$w_{0,g} = 2.5 \cdot 10^{-2}\,\text{mm} ,$$

[1]The standard configuration is the sum of all parameters for the modal register that are explained in this section.

an OQ of about 70 % is achieved. The model allows an individual adjustment of the anterior and the posterior attachment of the VF to the boundary. The values for the distance of one fold to the symmetry line can be varied between $-0.2\,\mathrm{mm}$ and $0.8\,\mathrm{mm}$. This makes it easy to model asymmetric glottal gaps and glottal leakage.

Spring stiffnesses

By observations of the VF movement a phase difference between the *mucosa* and the *vocalis* tissues was found. The *mucosa* masses follow the *vocalis* tissue with a phase lag of approximately 55°. This phase lag could be reproduced with the model with the following adjustments of the stiffnesses of the spring that couple *mucosa* and *vocalis* masses:

$$k_x^m = k_z^m = 2\,\mathrm{N\,m^{-1}}\ .$$

The factor for the nonlinear character of the springs is

$$\eta_{k,x}^m = \eta_{k,z}^m = 1 \cdot 10^{-4}\,\mathrm{m^{-2}}\ .$$

The spring stiffness that couples the *vocalis* mass to the *cartilago cricothyroideus* supports the active stress $T_{v,act}$ of the *vocalis* muscle and ensures an oscillation for small values of $T_{v,act}$. Measured values are not available, but

$$k_x^b = k_z^b = 3\,\mathrm{N\,m^{-1}}$$

has been proven to yield acceptable results for the standard configuration [Alh99].

The contact springs have not been measured yet, because of the impossibility to derive the deformation forces during contact of the VF *in vivo*. In fact, the contact springs have the task to model the deformation during the compression of the VF. The values for the contact forces have been first determined for the IF72 model and were used in most models since then:

$$\begin{aligned} k_{x,v}^c &= 3 \cdot k_x^b &&= 9\,\mathrm{N\,m^{-1}}\ , \\ k_{z,v}^c = k_{z,m}^c &= 3 \cdot k_z^m &&= 6\,\mathrm{N\,m^{-1}}\ , \\ k_{x,m}^c & &&= 0.125\,\mathrm{N\,m^{-1}}\ . \end{aligned}$$

Damping

The damping values are used according to [ST95]. For the open glottis the damping is

$$\begin{aligned} \xi_x^m &= 0.4\ , \\ \xi_z^m &= \xi_x^b = \xi_z^b = 0.2\ . \end{aligned}$$

For the closed glottis the damping is increased by 1 as assumed by [IK68].

$$\begin{aligned} \xi_x^m &= 1.4 \ , \\ \xi_z^m &= \xi_x^b = \xi_z^b = 1.2 \ . \end{aligned}$$

Active stress

The value of $T_{v,act}$ varies between 0 and 100 kPa after I. R. Titze [Tit73] but is assumed to be in the range 60..140 kPa after F. Alipour-Haghighi *et al.* [ATP89]. For the implemented model, a standard value of 30 kPa has been chosen.

Subglottal pressure

After G. Fant *et al.* [FHKL97] the subglottal pressure P_{sub} has values between 400 Pa and 800 Pa. However, P_{sub} can be significantly bigger when a loud voice is generated [MEV78]. With rising subglottal pressure, the amplitude rises, and the fundamental frequency of the voice signal increase slightly [BF94]. Most models use a value of

$$P_{lung} = 800\,\text{Pa}.$$

In this model, the lung pressure is not "switched on" but follows a ramp of the shape $P_{lung} = \frac{P_{lung}}{2}\left(1 + \tanh\left(\frac{n-30}{10}\right)\right)$ over 30 samples.

2.3.5 Summary of modifications

The differences to the 16-mass model of I. R. Titze [Tit73] and the body-cover model of B. Story and I. R. Titze [ST95] are the following:

- The coupling spring between *mucosa* mass and *vocalis* mass is replaced by a combination of a nonlinear spring and a damping element in $x-$ and $z-$direction for each segment.
- The coupling spring between *vocalis* mass and the boundary is replaced by a combination of a nonlinear spring and a damping element in $x-$ and $z-$direction for each segment.
- The exponential stress-strain curves of the *mucosa* and *vocalis* tissues in [Tit73] are replaced by a polynomial description [AT91].
- Jet generation and separation within the glottis is assumed. The position of the separation is chosen in dependence from the actual positions of the masses close to the smallest area within the glottis. The Bernoulli equation (2.1) is assumed between subglottal area and minimum glottal area. The diameter of the jet remains constant between the separation point and the supraglottal area. The pressure within the jet is assumed to be constant.

- Pressure recovery behind the separation point is assumed according to [IF72].
- The glottis model has been integrated into a waveguide model.

2.4 Simulations

Before the calculation of the mass movement starts, constants and variables that determine the boundary conditions for the equations are defined. Constants are e. g. air density ρ_0 and speed of sound at 37 °C. Variable, i. e. user defined, parameters are e. g. the lung pressure P_{lung}, the active stress on the *vocalis* $T_{v,act}$, the spring stiffnesses k and masses m of the oscillators. In literature usually a symmetric glottal gap is used due to the 1-dimensional DOF of the mass motion. The implemented model allows a more realistic configuration (e. g. triangular) of the glottal gap, as illustrated in the following simulations. The parameters can be changed during calculation and allow sample-wise modification. In chapter 6.3 simulation results of further configurations relevant for the singing voice synthesis are presented.

Figure 2.15 shows the setup of the graphical user interface (GUI) for the settings of the vocal fold model. One VF is shown from above, with the glottal gap situated to the left.

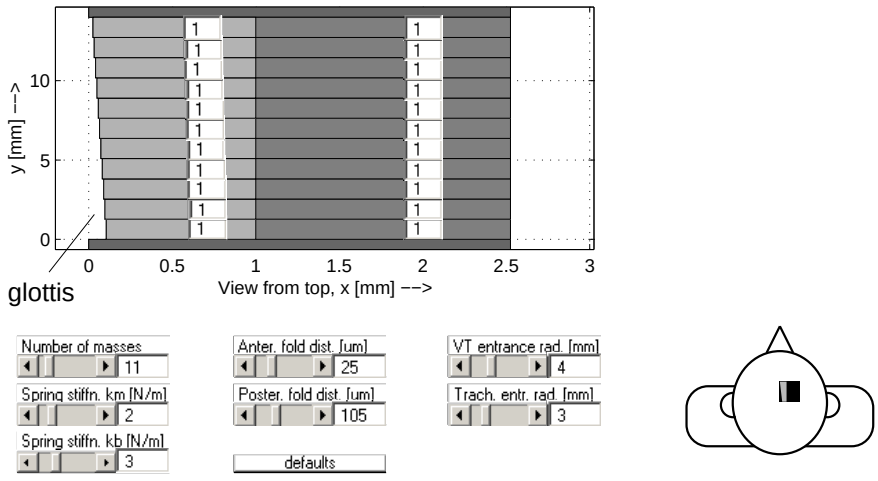
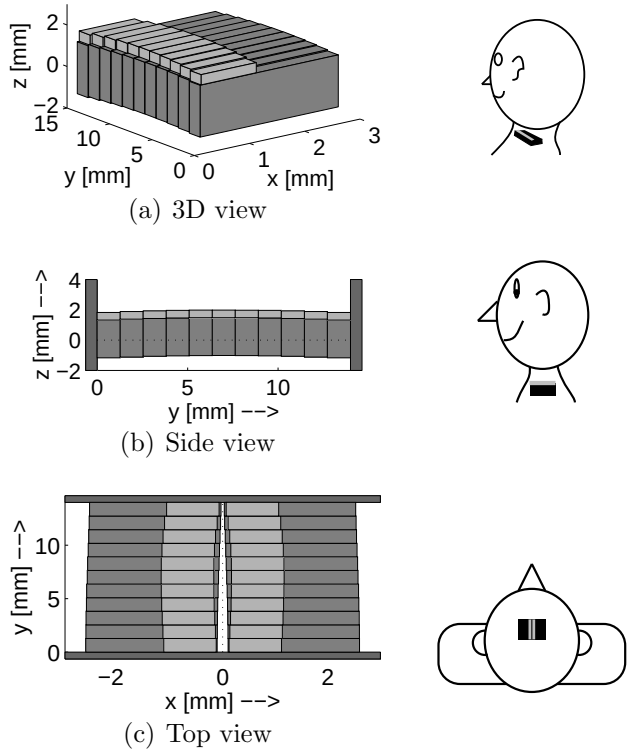


Figure 2.15: Parameter input (GUI) for the VF simulation (left) and position plus orientation of the model in the human body (right)

In this representation, the air flow passes the glottis in $z-$direction from below. The dark grey areas represent the *vocalis* masses whereas the light grey areas show the positions of the *mucosa* masses. On top of the masses, a number indicates the weighting factor of each mass. Below, the number of longitudinal masses, the spring

stiffness, the distances of the *anterior* and *posterior* rest position of the VF, and the entry radii of the sub- and supraglottal areas can be adjusted.

In Figure 2.16 the deflection of the vocal fold masses is shown for the above configuration of the modal register. In this visualization standard values for the VF have been chosen.



(a) 3D view

(b) Side view

(c) Top view

Figure 2.16: Deflection of the masses for modal register

Part (a) of the Figure presents a 3-dimensional view of the right VF model, seen from the rear left side. The middle part (b) of Figure 2.16 shows one vocal fold from the side. The left dark border represents the *cartilago cricoidea*, whereas the right dark border symbolizes the *cartilago thyroidea*. It can be seen that the subglottal pressure pushes the VF upwards. In part (c) both VF are seen from above. The phase offset between the *mucosa* masses and the *vocalis* masses during the closing of the VF can clearly be observed.[2]

Figure 2.17 shows the areas, the pressures, the flow, and the flow derivative during calculation. The picture at the top of the Figure 2.17 illustrates the constant values of the subglottal area A_{sub} (thin line) and the supraglottal area A_{sup} (dashed line),

[2]Animations of vocal fold movements for selected configurations can be found at the internet page `http://www.akustik.rwth-aachen.de/~malte/vox`.

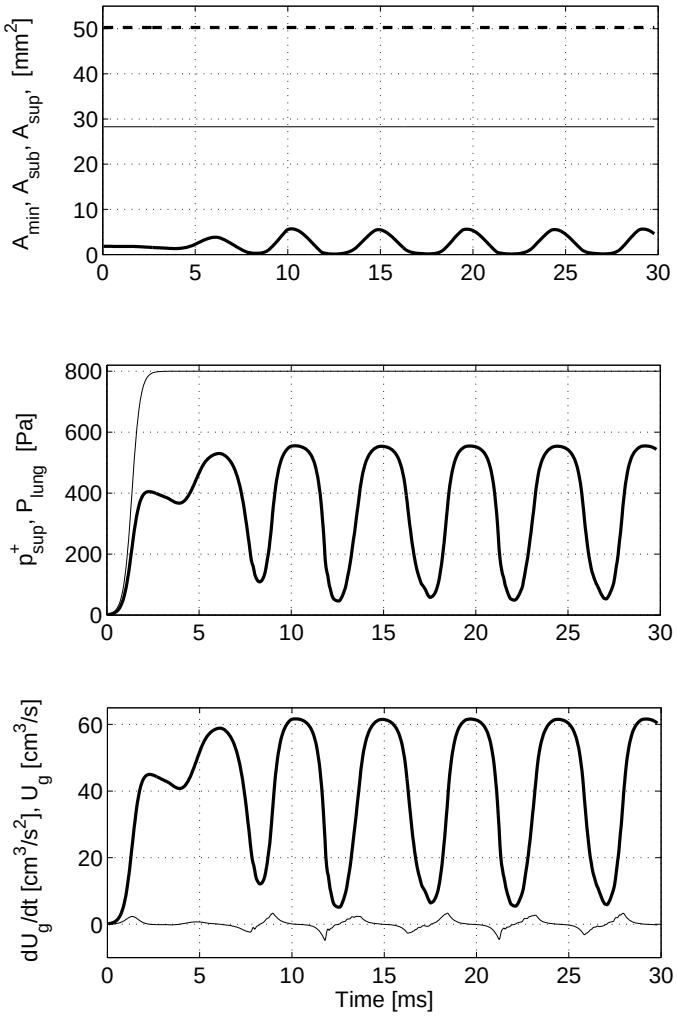


Figure 2.17: Result of modal register simulation: areas, pressures and flow

plus the variable minimum glottis area A_{min} (thick line). The picture in the middle visualises the rise of the lung pressure p_{lung} (thin line) and the development of the supraglottal pressure wave p^+_{sup}, (thick line, cf. eq. (2.51)). In the picture at the bottom of Figure 2.17 the flow (thick line) and the flow derivative (thin line) are shown.

The configuration was chosen so that the vocal folds did not close completely (set-up as in Figure 2.15). Further results of calculations with user defined parameter variations can be found in [Alh99].

In Figure 2.18 the forces on the *vocalis* masses are pictured. The forces on the *mucosa* are not shown since they are small compared to the *vocalis* forces. The

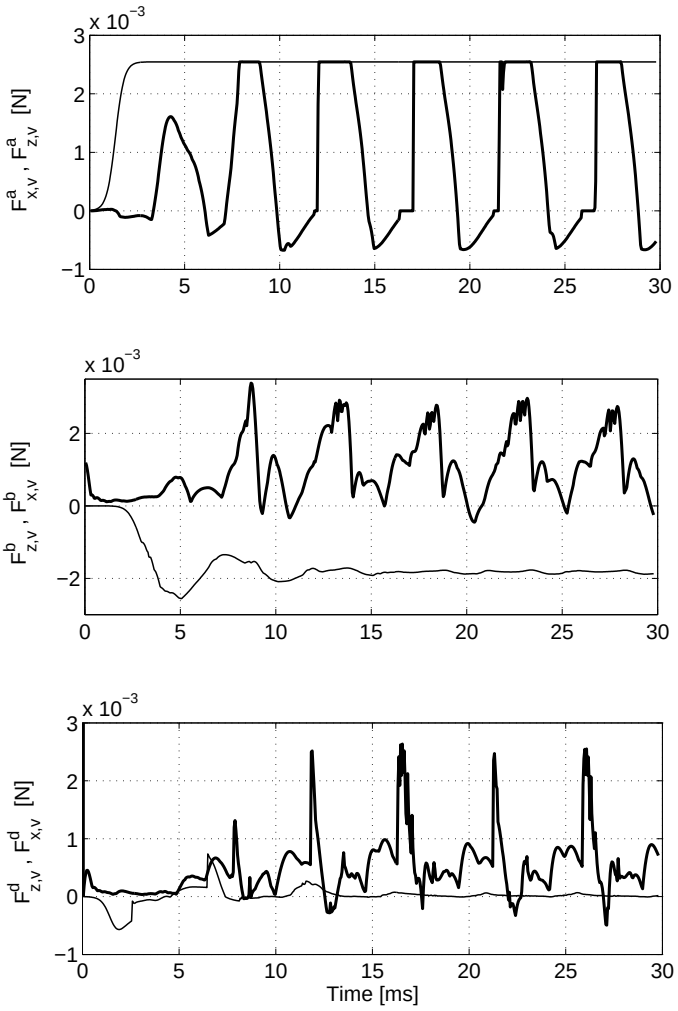


Figure 2.18: Result of modal register simulation: forces on the *vocalis* masses

upper picture displays the aerodynamic forces in $x-$direction (thick solid) and in $z-$direction (thin solid). It can be observed that the subglottal pressure pushes the *vocalis* masses continuously upwards ($z-$direction), and the Bernoulli force pulls them together in the open phase of the glottal cycle. The graph in the middle demonstrates the coupling forces, and the graph at the bottom of Figure 2.18 shows the damping forces that act on the vocalis masses.

2.5 Discussion

The better a model of the vocal folds imitates the behaviour of the human VF, the more complex the calculations become and the number of parameters increase drastically. The problem is to find a compromise between complexity and the applicability of a physical VF model. Real time calculations can be achieved only with simple models, but models that are capable of imitating e. g. voice pathologies require a certain complexity and more than two translational degrees of freedom.

The oscillation of a model is not a concern, as shown by the simple IF model or the symmetric model of N. Lous. Further interpretations must be done with care since the aerodynamic details of the VF oscillation cycle are very complex. One of the phenomena that remain to be modelled is the description of the pressure increase due to the closure of the vocal folds. It is expected that the rising pressure acts as an additional spring force on the VF that causes a delayed closure. This delay could reduce high frequency components in the glottis signal that are observed with most VF models. A further problem is the calculation of the jet separation point for complex VF geometries. The implemented model assumes a fixed separation point that is sufficient for self-sustained oscillation but simplifies the actual aerodynamics. More complex algorithms like those of N. Lous *et al.* [LHVH98] or X. Pelorson *et al.* [P$^+$94] should be adopted for the implemented mass model.

Modified two-mass models have been investigated for modeling of voice pathologies [Mer98, Dre01, HDS$^+$01, DA01]. There is certain interest in the medical field of communication disorders to apply the knowledge of voice generation to diagnosis and therapy of singers who suffer from pathologies such as edema or singer's nodules. The resulting pathological change in voice quality is generally called hoarseness. Acoustic measures for quantification of the pathologic voice change have been derived by many authors (e. g. F. Klingholz [Kli86], D. Michaelis *et al.* [MGS97]). A measure for pathologic laryngeal configurations called muscle tension patterns (MTPs) during singing has been introduced by Koufman [KRJ$^+$95]. Especially for the case of pathologic voice, some attempts are being made to relate extracted parameters from simple models to clinically relevant measures. It should be mentioned that a risk of physical modeling is the interpretation of results from an oversimplified model without sufficiently implemented physics behind it.

The present model aims at modeling complex VF movement with a minimum of parameters. The degree of complexity can be chosen by the number of segments along the length of the VF. In section 6.3 the application of the model to the synthesis of pathologic voices is presented.

The results of the simulations seem to match quite well the data published from authors who implemented similar algorithms [Tit73]. However, the results have not

yet been compared directly to measured supraglottal pressures or glottal flow values. The following paragraph describes a possible way to derive such values from direct measurements.

Measurements: For verification of the quality of the calculated VF signals a direct recording of the glottis sound pressure signal could be performed. In the following, such a procedure is described:

A miniature microphone (Sennheiser KE 4 211-4 [Sen00]) is put through one of the nostrils and placed close to the glottis. This procedure is done under supervision of a doctor due to the danger of a glottis cramp that can occur if an object touches the vocal folds.

Another microphone of the same type is placed close to the mouth at a fixed position relative to the head. When the pure glottis signal is to be determined, some problems have to be solved:

1. The signal is not independent of the surrounding structure. Reflections from the vocal tract superpose the generated sound field. Perceptually, the timbre of the glottis signal is not independent of the articulated phoneme.

   Two possibilities are proposed to compensate for the effect of the reflections from the vocal tract:

   - Damping of the reflected waves by application of e. g. damping tissue in the mouth cavity
   - Equalization of the input impedance of the vocal tract.

2. The sound pressure level inside the glottis is rather high (up to 140 dB). Thus, standard microphone types cannot be used.

3. To achieve only few interference of the measured sound field with the sound pressure transducer, the microphone should be as small as possible and placed just above the glottis under observation of a doctor using a mirror or an endoscope.

4. One problem with microphones in this humid environment is that the sensitivity drops drastically at frequencies above about 3000 Hz when the opening of the microphone capsule is covered by liquid. To protect the capsule against humidity it was covered by a shrinking tube that leaves about 5 mm space between membrane surface and the acoustic field. Liquid that blocks the opening of the tube can be removed by carefully blowing into the inlet of the capillary tube. This makes it possible to perform measurements without the need to extract and clean the microphones.

Chapter 3

Vocal tract

The vocal tract can be defined as the space downstream the glottis that ends with the mouth cavity *vestibulum oris* or the nostrils. During phonation for speech production the vocal tract geometry is changed intentionally to yield the target sound e. g. vowel or consonant. The geometry change is achieved by contraction or relaxation of several muscles that move tongue, lips, jaw, *velum* and *pharynx*. The transition between different sounds requires an instantaneous change of numerous control parameters. In this chapter, section 3.1 gives an overview of the geometry of the vocal tract. Section 3.2 presents models for the mathematical description of the vocal tract. In section 3.3 two implemented models are described. Section 3.4 contains results of simulations using these models. In the last section 3.5 measurement methods for acoustic properties of the vocal tract are presented.

3.1 Biomechanics

As a first rough estimate, the vocal tract can be described as a tube with space- and time-varying cross-section and variable length of about 14.5..17 cm for women, 17..20 cm for men and 7..10 cm for children [Sun87].

The nasal tract extends from the *velum* to the nostrils and is important for the generation of nasal sounds like [ŋ]. For vowels the nasal tract is not strongly coupled to the supraglottal volume and therefore the generation of nasal sounds will not be considered further on.

In Figure 3.1 an MRI plot of a singer's

Figure 3.1: MRI photo of a singer's head in the sagittal plane during phonation of [a:]

head is shown in the sagittal plane for configuration of the speech sound [a:]. The black cavity that extends from the narrow passage at the *larynx* to the lips is the vocal tract (VT). The most important muscle for geometrical changes of the vocal tract is the tongue. It allows fast variations of the vocal tract diameter in a huge range. Other relevant organs for modification of the VT are jaw, velum, teeth and lips.

3.2 Vocal tract models

For modeling of sound propagation through the vocal tract several approaches with different degrees of accuracy and computational effort are possible.

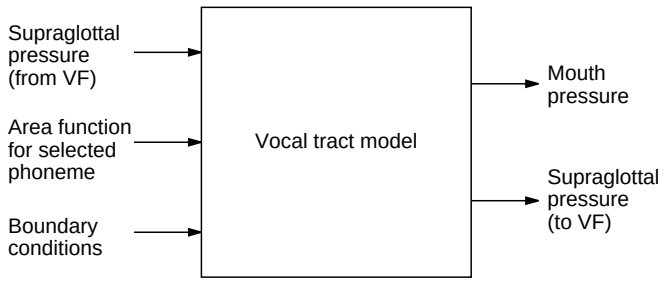


Figure 3.2: Function sketch of a vocal tract model

When the interaction between the vocal tract and the other functional components is considered, the schematic signal flow given in Figure 3.2 can be drawn. From the input parameters the supraglottal pressure is the most important. The pressure wave travels through the VT and is radiated at the mouth. The vocal tract area functions are determined by the articulatory constrains and vary in a broad range for different speech sounds [Krö98].

3.2.1 Finite element models

The most accurate approach for a description of the vocal tract would be the dynamic solution of the 3-dimensional wave equations, based upon a finite element mesh from 3-dimensional vocal tract data. Such an approach including aeroacoustic features such as losses and noise generation due to turbulences and vortices has not yet been realised. There are a number of reasons why this approach cannot be applied at present:

At first, the corpus of data needed for the construction of meshes for various speech sounds is still far from being complete. 3-dimensional MRI data from vocal tracts are difficult to obtain due to the time (and the money) needed for the 3D-scan of the head. Nevertheless, some effort has been made to build a database with such data [Krö00, KWMP00, ST96].

Secondly, a calculation based upon the finite element method (FEM) or finite difference method (FDM) must be carried out in the frequency domain. Therefore

it is only valid for stationary sounds. In the special case of singing voice synthesis the boundary conditions change rather slowly but despite this fact the calculational effort is quite high for an appropriate frequency resolution of the output signal.

FEM models can be used to calculate basic properties of the vocal tract. An example of this could be formant frequencies for stationary sounds such as long vowels. It has been shown by S. El-Masri *et al.* [EPSB98] by using the transmission line matrix (TLM) method, that the reduction of multidimensional wave propagation to one dimension can cause significant errors when calculating the vocal tract transfer function for frequencies above 5 kHz. However, damping of the voice signal for frequencies above 4 kHz is rather strong and the quality, i. e. the relative bandwidth, of the resonances at high frequencies is rather poor due to the increasing losses with increasing frequency (discussed in paragraph "losses" in the following section). An important point is that zeros that can be measured in the vocal tract transfer function [Pha95], cannot be simulated using a 1-dimensional waveguide model.

In general, the accuracy of a finite element calculation is limited by the uncertain estimation of the boundary conditions such as wall impedances, the exact form of the vocal tract, and the radiation impedance at the glottis and at the mouth. Another problem is the formulation of the aerodynamic equations in the vicinity of a complex, 3-dimensional and time-variant valve like the glottis.

The detailed information of a 3-dimensional mesh of the vocal tract is often reduced to a 1-dimensional acoustical model, the plane wave guide model.

3.2.2 Plane wave guide

The classical approach for modeling wave propagation in the vocal tract is the Kelly-Lochbaum (KL) model [KL62]. This type of model is also referred to as disc model, cylinder model or waveguide model. It assumes a 1-dimensional plane wave that emerges from the glottis end of the vocal tract and travels through a line of concentric cylinder segments with different cross-sections to the mouth or nose opening. Since the transitions between adjacent cylinder segments represent discontinuities in the area function, the number of segments must be sufficiently high for an accurate approximation of the real VT. The structure of the KL model is illustrated in Figure 3.3. The upper picture shows a sectional view of the cylinder segments, and the lower picture represents the signal flow chart for a waveguide model. The boxes with the letter R represent a multiplication with the reflection coefficient R, the letter T denotes the transmission factor T, and the expression z^{-1} represents a delay of the travel time between two subsequent segments. At each discontinuity between subsequent cylinder segments a part of the wave is reflected in the backward direction and one part is transmitted in the forward direction. At each time step and in each

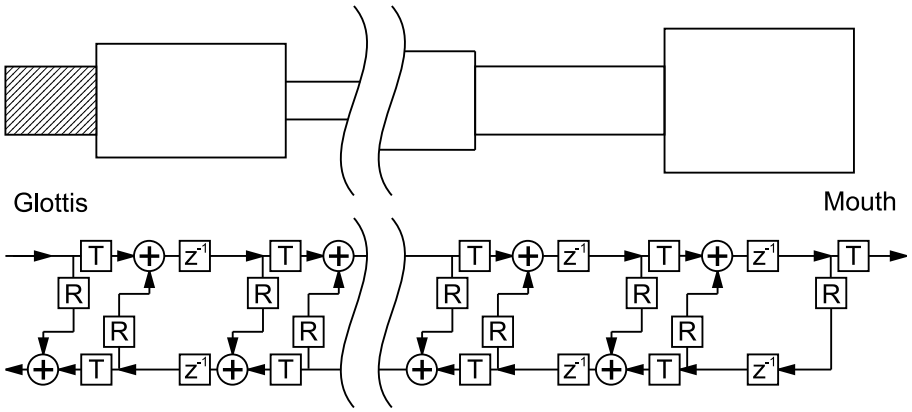


Figure 3.3: Schematic pressure flow in the waveguide model

segment, the resulting wave can therefore be regarded as a superposition of two waves that travel in opposite directions. A comprehensive study of the wave propagation in space and time has been done by J. Liljencrants [Lil85]. The following description is based on this work.

Static line model: At each junction of two segments, the two waves can be represented either as pressure waves P or flow waves U. Both representations can be converted one into the other using the equation

$$U = P\frac{A}{\rho c} \ . \tag{3.1}$$

A is the area of the cylinder segment, ρ is the density of air, and c is the speed of sound. The expression

$$Z = \frac{P}{U} = \frac{\rho c}{A} \tag{3.2}$$

is the acoustic impedance in the cylinder segments. In the following, the pressure representation will be used. However, the flow representation is possible as well and yields the same results with exception of differences for extreme parameter combinations (big flow through a very small passage) where the pressure representation can yield numerical instabilities [Lil85].

At each junction of two segments, numbered 1 and 2, the incident, reflected and transmitted pressure waves can be drawn as shown in Figure 3.4. In every segment at all times the pressure P is the sum of the forward wave p_+ (towards the mouth) and the backward wave p_- (towards the lungs):

$$P = p_+ + p_- \ . \tag{3.3}$$

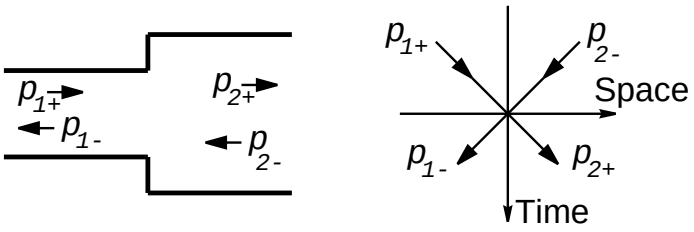


Figure 3.4: Scattering of waves at a junction of cylinder segments

On both sides of each discontinuity, the continuity of pressure and flow must be fulfilled. This leads to the following equation for the time-averaged pressure waves:

$$p_{1+} + p_{1-} = p_{2+} + p_{2-} \ . \tag{3.4}$$

A similar equation applies for the time-averaged flow waves:

$$\frac{p_{1+} - p_{1-}}{Z_1} = \frac{p_{2+} - p_{2-}}{Z_2} \ . \tag{3.5}$$

Rewriting (3.5) with (3.2) yields:

$$(p_{1+} - p_{1-})A_1 = (p_{2+} - p_{2-})A_2 \ . \tag{3.5b}$$

The combination of (3.4) and (3.5b) results in two equations for the unknown pressures on each side of the discontinuity:

$$\begin{aligned} p_{2+} &= \frac{2A_1 p_{1+} + (A_2 - A_1)p_{2-}}{A_1 + A_2} \quad \text{and} \\ p_{1-} &= \frac{(A_1 - A_2)p_{1+} + 2A_2 p_{2-}}{A_1 + A_2} \ . \end{aligned} \tag{3.6}$$

The definition of the reflection coefficient

$$R_{12} = \frac{A_1 - A_2}{A_1 + A_2} = \frac{Z_2 - Z_1}{Z_1 + Z_2} \tag{3.7}$$

allows the simplified form

$$\begin{aligned} p_{2+} &= (1 + R_{12}) \cdot p_{1+} - R_{12} \cdot p_{2-} \quad \text{and} \\ p_{1-} &= R_{12} \cdot p_{1+} + (1 - R_{12}) \cdot p_{2-} \end{aligned} \tag{3.8}$$

for the unknown pressures.

Losses: Only a relatively small part of the energy that is inserted into the VT is radiated at the mouth. From measurements that are described in section 3.5.1, a difference in sound pressure level (SPL) of 15..20 dB between signals recorded at the mouth and at the glottis was found. There are a number of ways on which energy gets lost during sound wave propagation through the VT:

- Wave propagation through the glottis towards the lungs
- Wave propagation into the VT walls
- Vibration of the VT walls due to acoustic excitation
- Heat conduction at the VT walls
- Dissipation due to laminar and turbulent flow

If the vocal tract is considered as an oscillator, the radiation of the sound at the mouth can also be called a loss of energy.

For losses that mainly depend on the flow rather than on the pressure, the energy loss can be expressed by a loss resistance Z_L that is in series with the cylinder segments (cf. Figure 3.5). The damping factor D^s can then be expressed by the ratio of Z_L to the sum of the adjacent impedances Z_1 and Z_2 in the cylinder segments:

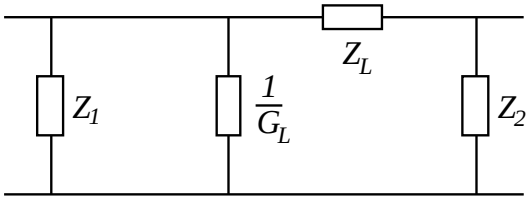


Figure 3.5: Electric circuit for losses

$$D^s = Z_L \frac{1}{Z_1 + Z_2} \text{ for serial losses.} \tag{3.9}$$

A similar approach holds for losses that depend on the pressure rather than on the flow. These losses can be represented by a shunt or parallel loss conductance G_L between the cylinder segments. A damping factor D^p can then be derived:

$$D^p = G_L \frac{Z_1 Z_2}{Z_1 + Z_2} \text{ for parallel losses.} \tag{3.10}$$

The lossy scattering equations for parallel losses are then

$$\begin{aligned} p_{2+} &= \left(1 + \tfrac{R_{12} - D^p}{1 + D^p}\right) p_{1+} - \tfrac{R_{12} + D^p}{1 + D^p}\, p_{2-} \quad \text{and} \\ p_{1-} &= \tfrac{R_{12} - D^p}{1 + D^p}\, p_{1+} + \left(1 - \tfrac{R_{12} + D^p}{1 + D^p}\right) p_{2-} \quad , \end{aligned} \tag{3.11}$$

and for the serial losses

$$\begin{aligned} p_{2+} &= \left(1 + \tfrac{R_{12} - D^s}{1 + D^s}\right) p_{1+} - \tfrac{R_{12} - D^s}{1 + D^s}\, p_{2-} \quad \text{and} \\ p_{1-} &= \tfrac{R_{12} + D^s}{1 + D^s}\, p_{1+} + \left(1 - \tfrac{R_{12} + D^s}{1 + D^s}\right) p_{2-} \quad . \end{aligned} \tag{3.12}$$

Note that the equations (3.11) and (3.12) only differ in two changes of sign in the weighting terms for the opposite wave components on each side of the discontinuity. In Table 3.1 the damping coefficients for these loss mechanisms are given.

Table 3.1: Losses in the vocal tract (after [Lil85])

Symbol	Formular	Standard value	Description
$D_g^p \cdot L$	$= \frac{\rho c U_g}{4AP_{lung}}$	$= 0.031$	Wave propagation through the glottis towards the lungs
D_{wvib}^p	$= \frac{R_0 \rho c}{4L_0^2 \pi^{3/2} f^2 \sqrt{A}}$	$= 537 \frac{1}{f^2\sqrt{A}}$	Vibration of the VT walls due to acoustic excitation
D_{wabs}^p	$= \frac{\rho c}{\rho_w c_w} \sqrt{\frac{\pi}{A}}$	$= 4.71 \cdot 10^{-4} \frac{1}{\sqrt{A}}$	Wave propagation into the VT walls
D_{heat}^p	$= \frac{\eta - 1}{\rho c^2} \sqrt{\frac{\lambda f}{\rho C_p A}}$	$= 1.610 \cdot 10^{-5} \sqrt{\frac{f}{A}}$	Heat conduction at the VT walls
D_{lami}^s	$= \frac{4\pi\mu}{\rho c A}$	$= 5.858 \cdot 10^{-7} \frac{1}{A\mathrm{m}}$	Dissipation due to laminar flow
D_{turb}^s	$= \frac{1}{8c} \sqrt[4]{\frac{\pi^{5/2} \mu U^3}{200 \rho A^{11/2}}}$	$= 1.234 \cdot 10^{-5} \frac{U^{0.75}}{A^{1.375}}$	Dissipation due to turbulent flow
D_{visc}^s	$= \frac{\pi}{c} \sqrt{\frac{f\mu}{\rho A}}$	$= 3.626 \cdot 10^{-5} \sqrt{\frac{f}{A}}$	Dissipation due to viscous flow
$D_{jet}^s \cdot L$	$= \frac{U}{2cA}$	$= 1.429 \cdot 10^{-3} \frac{U}{A}$	Dissipation due to jet generation (turbulence and noise)
$D_{lip}^p \cdot L$	$= \frac{\pi f^2 A K_s}{2c^2}$	$= 1.795 \cdot 10^{-5} f^2 A$	Sound radiation at the lips

The following symbols have been used[1]: density of air ρ, density of the VT wall $\rho_w = 1000\ \mathrm{kg\,m^{-3}}$, viscosity of air μ, speed of sound in air c and in the VT wall $c_w = 1500\ \mathrm{m\,s^{-1}}$, glottal flow U_g, VT area A, lung pressure P_g, VT wall surface resistance $R_0 = 12 \cdot 10^3\ \mathrm{N\,s\,m^{-3}}$, specific surface inductance $L_0 = 20\ \mathrm{kg\,m^{-2}}$, adiabatic constant η, viscosity of air μ, thermal conductivity λ, and the frequency-dependent radiation impedance R_s (cf. section 5.1).

Extended plane wave guide models: Quite a number of publications deal with modification and extensions of the KL model.

In case of fast varying VT configurations, the scattering equations (3.8) and (3.11-3.12) are no longer valid. Two different solutions for this problem have been published by S. Maeda [Mae77] and H. W. Strube [Str82]. A discussion of the advantages of each approach can be found in [Lil85] and [Str00]. However, for the task of modeling

[1] Standard values for the symbols can be found in appendix A.

sustained vowels, the problem of time-variant VT area functions is not crucial and will not be considered further.

A drawback of the KL model is the limitation of the sampling frequency to the inverse travel time between two elements. As a consequence, the whole vocal tract model has to be run at this sampling rate, which is much higher than necessary for voice synthesis. This problem has been addressed by V. Välimäki [Väl95], who developed an algorithm for the time-discrete but space-continuous calculation of waveguide models using fractional delay filters.

Another problem is that the segments must be equally spaced which corresponds to a local oversampling for parts of the vocal tract that do not change significantly in geometry.

Recently, H. W. Strube [Str00] showed that the approach of the KL model can be extended to take into account variable vocal tract lengths during calculation.

As an alternative to the discretisation of the vocal tract by equally spaced discs, a recent approach using cone segments is proposed.

3.2.3 Multiconvolution

The continuous time-interpolated multiconvolution (CTIM) algorithm of A. Barjau *et al.* [BKC99] is an improved multiconvolution technique based on the work of J. Martinez *et al.* [MA88, MAC88]. The method described in these papers uses reflection functions to describe discontinuities in musical instruments, especially woodwinds.

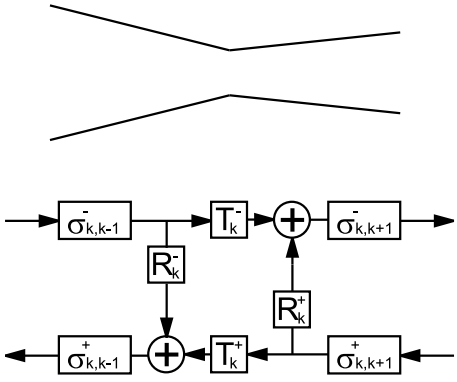


Figure 3.6: Method scheme of the CTIM algorithm

In the CTIM or cone model wave propagation is modeled by two opposite waves as in the KL model, but the propagation is calculated locally at both sides of each discontinuity (cf. Figure 3.6).

An integral method is used for the convolution of an incident pressure wave p with the time-continuous propagation function σ:

$$\sigma^{\pm}\left(t-\frac{L}{c}\right) = B^{\pm}\varepsilon\left(t-\frac{L}{c}\right)\cdot\frac{\xi}{2\sqrt{\pi}(t-L/c)^{1.5}}\,e^{\frac{-\xi^2}{4(t-L/c)}}\ . \tag{3.13}$$

In this equation L is the distance between two discontinuities, ε is the Heaviside unit step function, B^+ is the ratio $\frac{r_2}{r_1}$ and B^- the ratio $\frac{r_1}{r_2}$. ξ represents the damping coefficient that has been defined by C. Nederveen [Ned69]:

$$\xi = \sqrt{2}\zeta_0\frac{L}{c^{1.5}D_m}\ . \tag{3.14}$$

In equation (3.14) the factor ζ depends on ambient conditions [BKC99] and has been given a value of 0.135 [MAC88]. D_m is the average diameter of two segments.

The reflection and transmission of waves at discontinuities is modeled with reflection and transmission functions $F(t)$ of the form

$$F(t) = a_0\delta(t) + a_1 e^{b_1 t} + a_2 e^{b_2 t} + a_{31} e^{b_3 t} + a_{32} t e^{b_3 t} \ . \tag{3.15}$$

Due to the time-continuous calculation of the propagation functions, the sample rate F_s can be chosen arbitrarily. For different discontinuities as shown in Figure 3.7, the

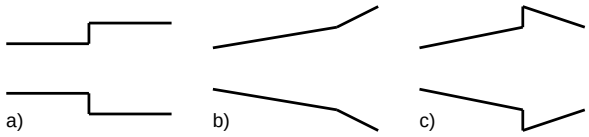


Figure 3.7: Segments used by the multiconvolution algorithm CTIM

values of a and b depend on the geometry. This calculation is not dependent on the sampling rate and only needs to be performed at each discontinuity. In many cases the vocal tract geometry can be represented by a small numbers of cone segments as described in section 6.3.

3.3 Implemented models

Two models have been implemented: a waveguide and a multiconvolution model.

3.3.1 Reflection type line analog

The implemented waveguide model is based upon the general structure of the iterative approach by Barjau [BKC99]. Since only the basic features necessary for plane wave propagation in cylindrical segments were implemented, the final algorithm corresponds to the description given in section 3.2.2. Of the damping factors described in Table 3.1 only the frequency-independent losses are taken into account by multiplication of the transmitted pressure wave by a damping factor. For the frequency-dependent sound radiation at the mouth two different models have been used:

1. the reflection function corresponding to the radiation impedance of a oscillating sphere (cf. equation (5.2) in chapter 5), and

2. a radiation impedance that models the radiation impedance of a piston in a baffle.

In Figure 3.8 the signal flow for the waveguide calculation is shown. For each sample, the user-defined parameters are imported from the main program *VOX*. Depending on the spacing of the vocal tract data, the sample rate is calculated (see equation (3.16) on page 54). The reflection coefficients are calculated using equation (3.7). When the noise generation module is active, the sound pressure in appropriate segments is augmented by turbulent noise. Now the vectors are shifted and the loop is repeated until the time limit is reached. The algorithm can be used in interactive mode for the generation of single pressure samples or in stand-alone mode for the calculation of vocal tract transfer functions (VTTF). A detailed description of the algorithm can be found in [Rei99]. The MATLAB code of the algorithm is given in appendix E.

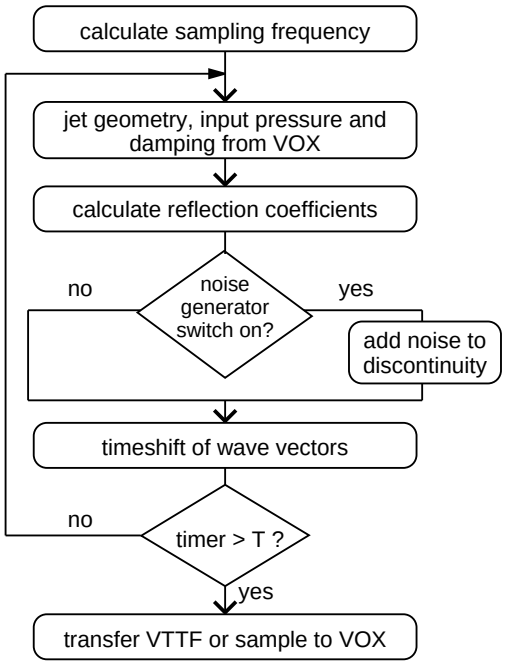


Figure 3.8: Algorithm of the waveguide model

3.3.2 Multiconvolution technique

The algorithm of A. Barjau *et al.* [BKC99] is an improved multiconvolution technique based on the work of J. Martínez *et al.* [MA88, MAC88]. Whereas the time-domain algorithm described by Martínez requires a spacing of the discontinuities that is restricted to multiples of phase velocity times time step, the interesting aspect of the CTIM (continuous-time interpolated multiconvolution) algorithm is its independence of the spatial VT discretisation and the sample rate. The flowchart of the multiconvolution algorithm is shown in Figure 3.9. Originally, the algorithm has been invented for the impedance calculation of woodwind instruments with toneholes and moderately changing cross-sections. Therefore, the transmission of the generated sound waves through the mouth opening is not implemented. Moreover, in contrast to the waveguide algorithm no noise insertion has

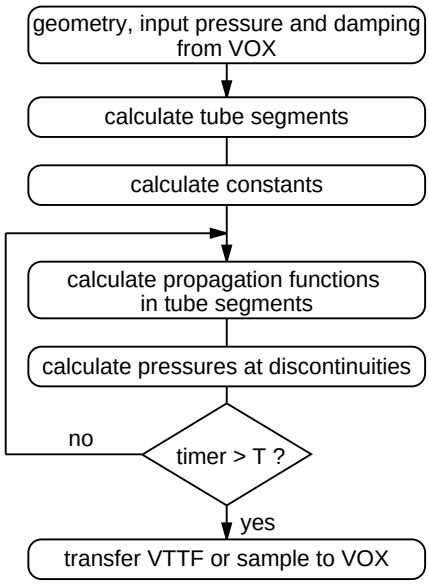


Figure 3.9: Algorithm of the multiconvolution model

yet been implemented. The calculation scheme differs from the waveguide approach in the following aspect: An update of the tube segment geometry is necessary after changes in the vocal tract geometry only. Therefore, the constant parameters that are used for the convolution routine can be calculated once before the loop is entered. As a consequence, the algorithm is much faster than the waveguide approach.

The calculation of the tube segments is done with the help of a graphical user interface (cf. Figure G.2 in the appendix, and the following section) and allows the instantaneous interpretation of the results during construction. A pseudo-MATLAB code of the algorithm is given in appendix F.

3.3.3 Parameters

For both implemented models only a few relevant parameters exist: equivalent area functions, sample rate and damping.

Equivalent area functions: The configuration of the geometrical parameters of the model is done by a GUI that allows the choice of settings for different vowels and modifications of the default configuration.

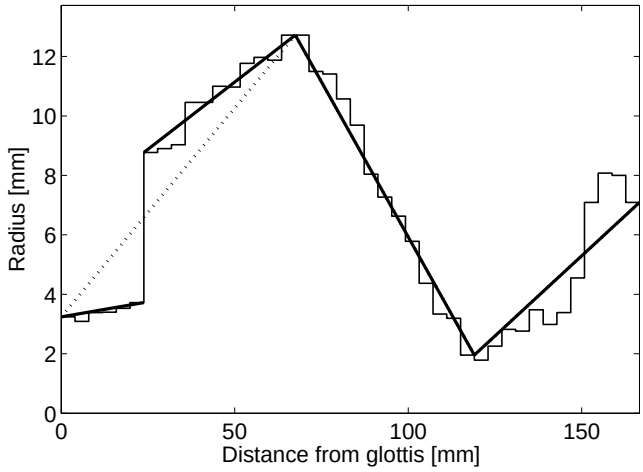


Figure 3.10: Construction of the cone segments, vowel [i:]

In case of the waveguide model, the data for the equivalent area functions (EAF) have been taken from B. Story and I. R. Titze [ST96] for 10 American vowels, from B. J. Kröger *et al.* [KWMP00] for 6 long German vowels and from S. Adachi and M. Yamada [AY99] for 4 overtone sounds.

For the multiconvolution model, the cone segments are directly calculated from the raw EAF data with the following algorithm:

- Connect left side of the first and right side of the last segment by a line
- Calculate the location of the maximum distance between the EAF and the line
- Split the line in two lines at the position of the maximum deviation and connect the open ends with the right or left side of the EAF segment at the location of maximum distance

- Repeat the above steps for each new cone segment until the desired maximum number of segments is reached

If a sharp discontinuity is encountered, i. e. a significant difference in area between two segments, the algorithm chooses different area values for each side of the discontinuity. Consequently, the number of segments can be reduced in such cases.

In Figure 3.10 the result of the algorithm for five lines, i. e. four segments is shown. The thin solid line indicates the raw EAF, the bold solid line represents the radii of the cone segments. The dotted line shows the course of the algorithm by a cone segment prior to being splitted in two lines across a strong discontinuity. In contrast to linear interpolation that aims at minimising the sum of deviations from the given EAF for each segment, the above algorithm might be less accurate but ensures that the minimum and maximum values never exceed the true EAF values. This has been found to be important especially for EAF with narrow passages.

The mode of calculation (cylinder or cone model) can be chosen in a popup-menu of the GUI as shown in Figure G.2 in appendix G.

Sample rate: A difference of the two implemented models is the control of the sample rate. Whereas the sample rate F_s can be chosen arbitrarily in the cone model (cf. section 3.2.3), in the waveguide model the sample rate is determined by the distance of the cylinder segments.

For the data of [ST96], F_s is calculated as the ratio of the sound speed and the distance between two subsequent segments:

$$F_s = \frac{c_0}{\Delta z} = \frac{350\,\mathrm{m\,s^{-1}}}{3.9683\,\mathrm{mm}} = 88\,200\,\mathrm{Hz}. \tag{3.16}$$

This value is significantly higher than necessary for voice generation.

Damping: Also the algorithm for the calculation of the damping is different in both models.

In the waveguide model the transmitted amplitude is multiplied by a factor $1-\xi$ at each discontinuity and the value for the damping of the pressure waves is $\xi = 0.005135$.

In the cone model the damping is assumed to be $\xi = 0.135$ [MA88]. In this model, ξ is an exponential factor that takes into account viscous losses.

3.4 Simulations

The vocal tract transfer function (VTTF) is the most expressive description of the acoustic properties of the VT. However, for numerical evaluation a time-domain representation is more appropriate. The sound pressure impulse response of an acoustic

system is a unique measure that relates the sound pressure at its output end (here: mouth) to a defined input signal at its input (here: vocal folds). For sound generation using the classical oscillator-filter model the VT impulse response is convolved with the sound pressure signal at the glottis yielding the sound pressure at the mouth.

In waveguide models the impulse response is not calculated explicit for sustained sounds since the essential interaction of vocal tract and vocal folds (cf. section 6.2) would require a time-variant impulse response calculation. Instead, an iterative algorithm is used for a sample-wise calculation of the sound pressure between the segments. The calculation of the impulse response is performed indirectly: as input signal a vector [1 0 0 0 0 ...] consisting of a Dirac impulse with trailing zeros is fed into the VT. For each sample, the actual pressures of the two pressure waves propagating back and forth in the VT are stored at all discontinuities. In Figure 3.11 the propagation of the sound wave in the vocal tract between glottis (bottom) and mouth (top) is depicted for the vowel [a:]. The travelling time of the sound waves through the vocal tract can be read from the plot (about 0.5 ms) and the locations of sharp discontinuities (high amplitudes). The amplitude of the pressure signal at the mouth is very weak compared to the glottal pressure. Both algorithms were evaluated with Dirac impulse excitation and the stand-alone calculation mode. A comparison of both methods is given in Figure 3.12 for the configuration corresponding to the vowel [a:].

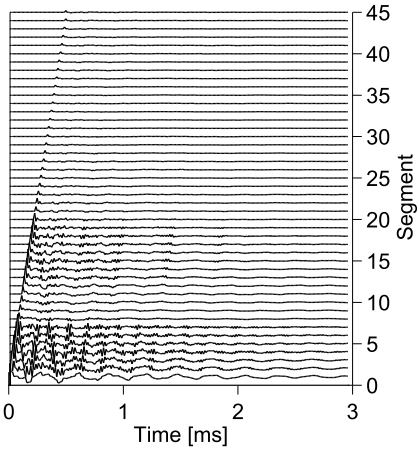


Figure 3.11: Waterfall diagram of sound propagation in the vocal tract, sound [a:]

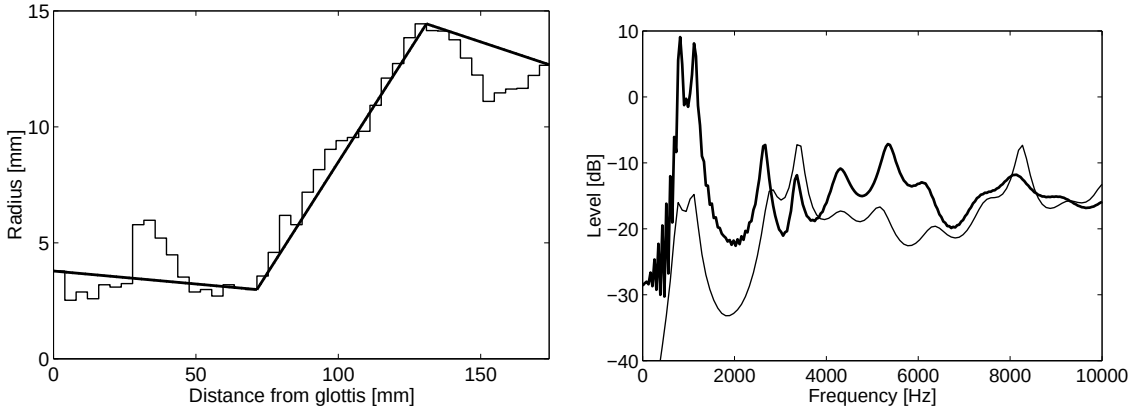


Figure 3.12: Comparison of VT parametrisation (left) and calculated VTTF (right) using the disc model (thin line) and the cone model (thick line), sound [a:]

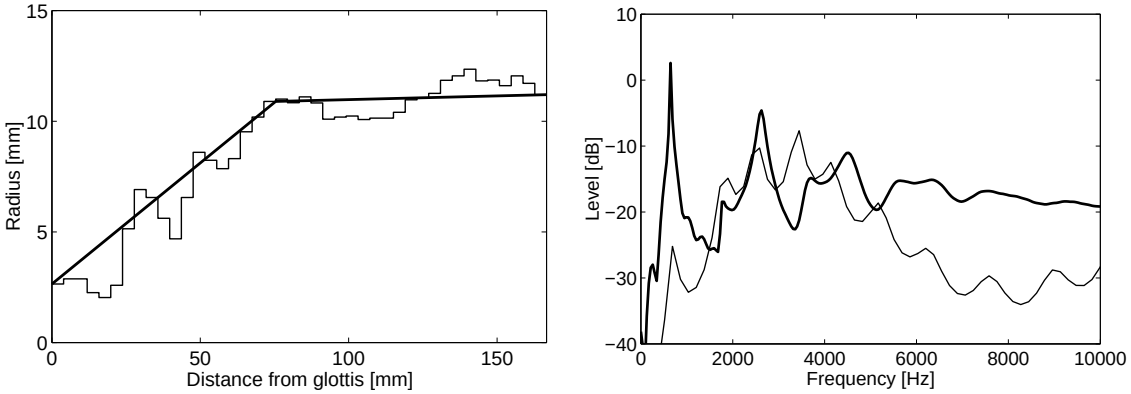


Figure 3.13: Comparison of VT parametrisation (left) and calculated VTTF (right) using the disc model (thin line) and the cone model (thick line), sound [æ:]

.

With both algorithms the correct formant frequencies are obtained (cf. Table B.2 in appendix B). However, the frequency distribution differs significantly: a difference of ca. 25 dB can be observed for the formants below 3 kHz.

The shape of the curve obtained from the waveguide calculation is in good agreement to measurements [Pha95]. The deviation between both methods thus could be explained by the missing implementation of the frequency-dependent radiation function in the multiconvolution model (cf. section 5.1). The VTTF of other sounds also exhibits a rather constant deviation of 25 dB for low frequencies (cf. Figure 3.13).

An additional problem occurred while calculating the VTTF of convergent EAF: the algorithm was not stable if either a huge number of segments or strongly convergent segments were used. In a recent review J. O. Smith mentioned the problem of using convergent cone segments for waveguide calculations [Smi96]. Though the algorithm was not applicable for the simulation of sustained phonation. Nevertheless, the calculation of the formant frequencies using the multiconvolution model has been proven to yield satisfactory results for the estimation of formant frequencies and bandwidths. Within limits, the stability of the algorithm could be improved by choosing a moderately higher damping value. Nevertheless, the calculation time was restricted to about 10 ms before instabilities occurred.

3.5 Measurements

So far the characterization of the acoustic vocal tract properties has been investigated in detail using indirect methods [FL71, Pha95]. In this section two novel methods are presented. The first method directly measures the vocal tract transfer function between glottis and mouth (VTTF), whereas the other method determines the vocal tract impedance at the mouth (VTMI). Both methods function as tools for verification of the simulation results from the two implemented methods, the waveguide and the multiconvolution technique. In literature VTTF and VTMI measurements have been proven to be useful for acoustic characterization of the vocal tract (e. g. [Pha95, DSW96]). The description of the novel VTMI method in section 3.5.2 is an extended version of an article that has recently been accepted for publication [KN02].

3.5.1 Vocal tract transfer function

The direct measurement of the vocal tract transfer function (VTTF) can be performed by a two channel measurement of the sound pressures at the glottis and in front of the mouth. In general, two different methods can be used: external excitation of the vocal tract and internal excitation. Both methods will be described in the following section.

Depending on the kind of excitation, the ratio of the spectra of both signals yields the transfer function

$$\begin{aligned} H_{glottis\rightarrow mouth}(f) &= \frac{\mathrm{FFT}(s_m(t))}{\mathrm{FFT}(s_g(t))} \quad \text{for internal excitation, or} \\ H_{mouth\rightarrow glottis}(f) &= \frac{\mathrm{FFT}(s_g(t))}{\mathrm{FFT}(s_m(t))} \quad \text{for external excitation.} \end{aligned} \tag{3.17}$$

In the equations, s denotes a signal that corresponds to the impulse response of a system measured at the mouth (index m) or at the glottis (index g). FFT stands for fast Fourier transform. Measurements of the VTTF have been performed both on vocal tract models and *in vivo*.

In vivo measurements – Internal excitation

When the vocal fold signal is used as a sound source, care must be taken that the excitation of the vocal tract is performed in a way that provides a sufficient supply of energy to the formants. If the subject utters a sustained sound with approximately constant fundamental frequency, this prerequisite is not fulfilled. The solution is the production of either a broadband noise such as whispering or a voiced sound with varying fundamental frequency over a range of one octave during the measurement period. Thus it is ensured that energy is fed in all frequency bands from the fundamental upwards.

For determination of the transfer function the following steps have to be performed:

1. Clean both microphones
2. Insert the first miniature microphone (type KE 4-211-2) through the nose down to the glottis
3. Place the second microphone in front of the mouth
4. Sing and record loud [a:] with loud voice, check for clipping in both channels
5. Sing and record the sound [a:] swept from lowest possible to highest note in modal register (ambitus > 1 octave)
6. Repeat step 5 for the vowels [e:], [i:], [o:], [u:], [æ:], and [ɔ:]
7. Carefully pull out the internal microphone and clean it
8. Backup data!

The procedure must be supervised by a phoniatrician using a mirror to ensure a proper positioning of the microphone just above the glottis.

It should be noted that the manoeuvre is rather unpleasant and might not be accepted by people, who could not bear the microphone and cable in their *pharynx*. Another problem is the danger of a vocal fold cramp that can occur if the microphone touches the vocal folds. Therefore, the *in vivo* measurements described in this section have been carried out for only one subject (male, age 33).

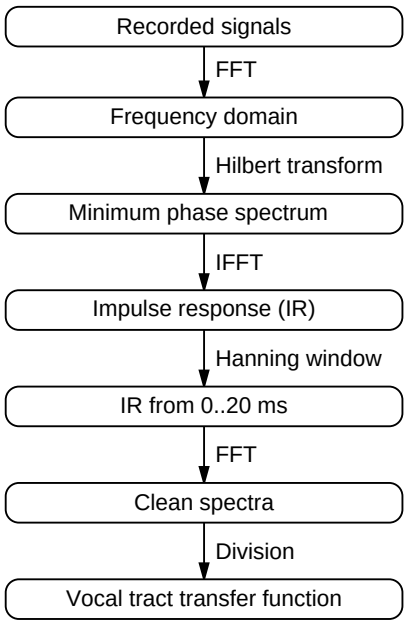
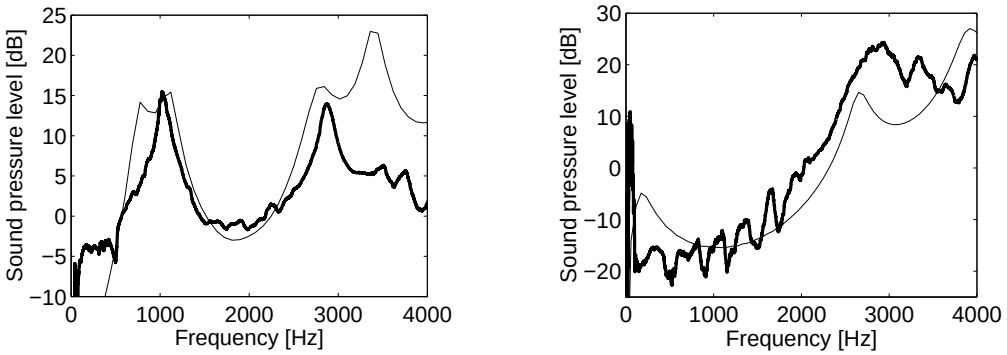


Figure 3.14: Signal flow of VTTF postprocessing

The 2-channel recordings are postprocessed as described in Figure 3.14. The application of the window in the time domain removes unwanted noise possibly generated by reflections outside the vocal tract.

In the following, the measurement results are compared to simulations, although the EAF data used for the simulations is certainly not identical to the vocal tract area functions of the subject under test. Nevertheless, the formant frequencies should be similar in both cases and, within a limited region, independent of the

subject (cf. appendix B.2). A comparison of the results from a measured VTTF with a simulated VTTF is shown in Figure 3.15 for the vowels [a:] and [i:]. Further measurements have been performed but the poor signal-to-noise ratio (SNR) excluded the results from an evaluation. Between 500 Hz and 3000 Hz a good agreement of



(a) vowel [a:] (b) vowel [i:]

Figure 3.15: Comparison of simulated (thin) and measured VTTF (thick solid) using internal excitation

simulation and measured VTTF can be found. However, the double formants at 1000 Hz and 3000 Hz are not separated in the measured curve. A similar finding for the vowel [y:] has been reported in [FL71]. For frequencies below and above this range, the formants cannot be clearly identified in the measured curves. For the high frequencies this problem might be caused by the high damping of the signal of the glottis microphone for frequencies above 3 kHz due to a possible liquid closure of the small entry channel in front of the microphone membrane.

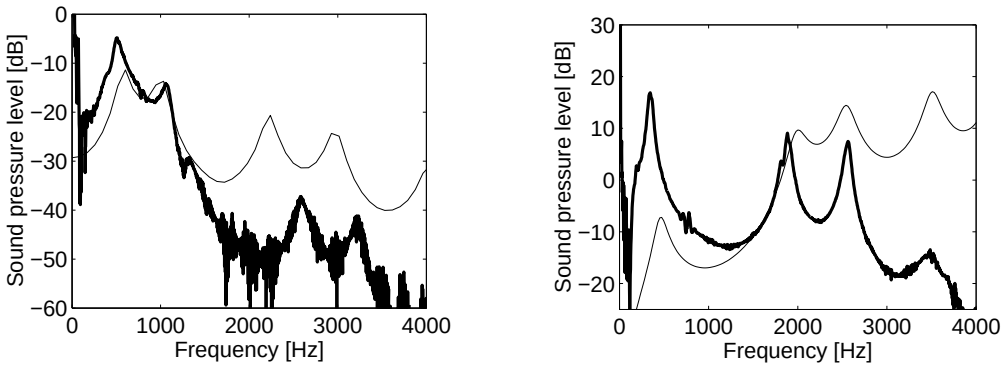
In vivo measurements – External excitation

Sweep signals have already been used for the investigation of the vocal tract configuration by O. Fujimura *et al.* [FL71]. A method has been described by Y. Pham Thi Ngoc [Pha95] for the case of an externally induced vibration of the larynx by application of a force. Another, non-invasive way is the excitation of the vocal tract with a broad-band signal at the mouth. The new aspect in this work is the acoustic excitation using a sweep correlation technique for enhanced dynamics.

In the following, a method is described that uses a swept sine wave of length 2^n, $n = 10..14$ as excitation signal at the mouth. The procedure for signal processing is very similar to the one described in the previous paragraph (i. e. internal excitation). The main difference is the choice of the excitation point: here, a sweep signal is generated at the opening of the mouth cavity. As a consequence, the signals are

processed similar to the case of internal excitation, apart from the final step where the ratio is the glottis signal divided by the mouth signal. Since this method does not require a voice signal, the SNR can be improved if no phonation takes place during measurement. The articulation can be kept constant if the subject at first utters the desired speech sound and then tries to keep the articulation configuration in mutism.

A comparison of the results from a measured with a simulated VTTF using the waveguide algorithm is presented in Figure 3.16 for the English vowels [ɔ:] and [I:]. The comparison of the measurements with the simulated VTTF exhibits a general



(a) vowel [ɔ:] (b) vowel [I:]

Figure 3.16: Comparison of simulated (thin) and measured VTTF (thick solid) using external excitation

agreement of the formant frequencies. However, the relative and absolute levels of the formants do not match at all. A reason for the low level of the measured higher formants could be the problematic positioning of the microphone in the larynx. When the opening of the microphone happens to touch a VT wall, a significant loss of sensitivity occurs for high frequencies ($> 2..3\,\mathrm{kHz}$). Another reason for the level differences might be due to the assumption of frequency-independent propagation losses in the waveguide simulation.

Measurements on models

For verification of the *in vivo* measurements and simulations of phonemes, experiments have been performed with two different types of physical 'hardware' models.

One model is a rigid model with a fixed, continuous equivalent area function. The other model is assembled from a number of concentric PVC discs with diameters that are chosen to approximate a desired equivalent area function.

Aluminium model: The rigid model is a 1:1 aluminium model consisting of two half plates that are milled to form a vocal tract shape that follows the equivalent area function for the phoneme [Λ:]. Two variations have been produced: one with a bent tube and one with a straight tube. Both models were coated with a thin silicon layer that imitates the surface impedance of the skin for high frequencies and introduces laminar and viscous losses. Figure 3.17 presents the half plates of both models. For

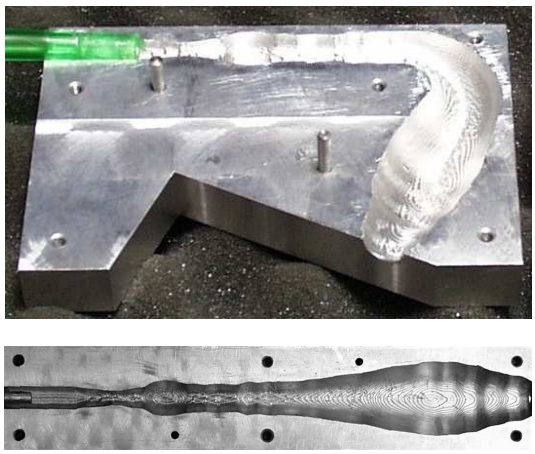


Figure 3.17: Bent (top) and straight (bottom) aluminium model

the bent model (top) in the upper left edge the tube can be seen that serves as a waveguide of the excitation signal into the vocal tract. Below the bent model, the corresponding straight model is pictured. The VTTFs between 0 Hz and 8 kHz for both models are visualised in Figure 3.18.

The frequency plot shows a good agreement of the frequencies and bandwidths of the resonances for frequencies below 5 kHz. At 5.8 kHz a shift between both resonance curves can be observed, although the deviation in amplitude does not exceed 3 dB. It can be concluded that for frequencies below 5 Hz the assumption of a straight instead of a bent VT is justified. The deviation at higher frequencies is in agreement with calculations using 3-dimensional wave propagation approaches [EPSB98].

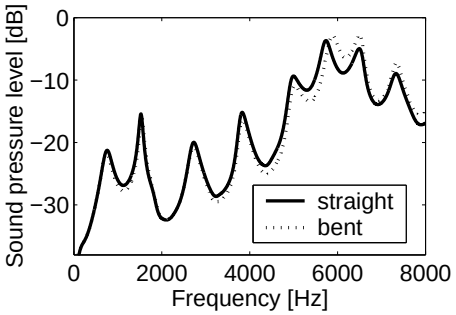


Figure 3.18: Comparison of the VTTFs of straight and bent aluminium model

Washer model: The second model can be called a "washer" model, i. e. a straight model that consists of a number of discs with variable cross section. The modular assembly concept allows the modeling of

vocal tract configurations of variable total length and arbitrary EAF. The thickness of the discs ($\approx 4\,\mathrm{mm}$) has been chosen to match the spacing of the MRI data of Story and Titze [ST96]. The set-up of the model is visualised in Figure 3.19. Discs

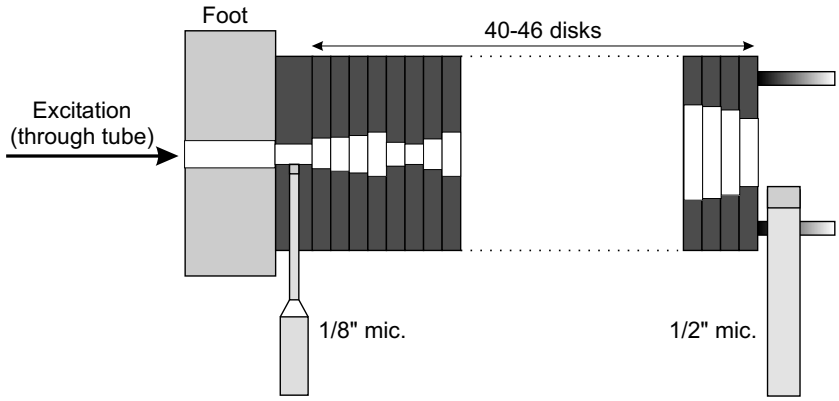


Figure 3.19: Set-up of the washer model

of areas that match the EAF of the straight aluminium model were chosen, lined up on two screws and fixed with wing nuts. The microphones were fixed at positions corresponding to the set-up of the aluminium model and equivalent measurements of the VTTF have been carried out with the washer model. In Figure 3.20 the result of a measurement on the washer model in the configuration of the aluminium model is shown. The frequencies of the resonances are closely related but not identical for frequencies below ca. 3 kHz.

For frequencies over 1 kHz the bandwidth of the resonances measured in the washer model is significantly higher, corresponding to a higher damping. The losses that could cause this frequency-dependent behaviour are heat conduction into or flow dissipation due to viscous flow. Since a comparison of measurements with and without coating did not exhibit a difference in the resonance bandwidth [Rei99], viscous flow due to the discontinuous surface structure of the washer model seems to be the reason for the relatively high damping of the higher frequencies.

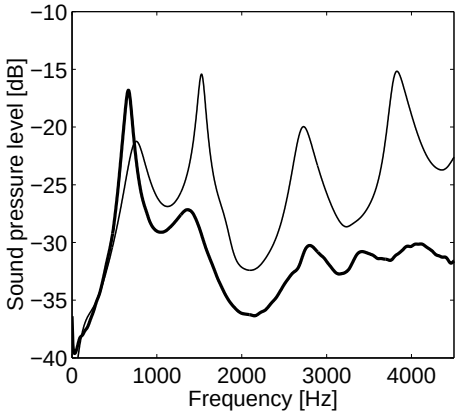


Figure 3.20: Comparison of the VTTFs of washer (bold) and aluminium model (thin)

As a conclusion from the above measurements the implementation of viscous losses are important for physical modeling of the vocal tract.

3.5.2 Vocal tract impedance

Phonemic speech information transmitted by acoustical signals is encoded by the time-dependent acoustic response of the vocal tract that is formed by different articulatory organs such as lips, tongue, *velum* and *larynx*. In order to define the phonemic positioning of these articulatory organs, X-Ray imaging techniques, MRI and ultrasonography [SSH+81, WBH+90] have been applied. However, some of those procedures are very costly. Alternative acoustical approaches are available to estimate the vocal tract transfer function (VTTF) [FL71] or even the vocal tract shape [SG71], although in an indirect manner. The vocal tract impedance at the mouth (VTMI) presented in this paper is a characteristic descriptive measure of vocal tract acoustics as is the vocal tract transfer function (cf. [Pha95]). Both measures can be used to derive the frequencies and bandwidths of the resonances of the vocal tract. Whereas a direct determination of the VTTF requires an in-situ measurement of the glottal sound pressure, the impedance method is non-invasive. Indeed the vocal tract impedance measured at the mouth has been used previously to characterise the resonant frequencies of the vocal tract [ESW97], and to provide feedback on the vocal tract configuration as a training aid for the correct pronunciation of vowels [DSW96, DSW97].

Method

The VTMI method is based upon the set-up described in [ESW97], which consists of a loudspeaker being attached to an impedance matching horn with an inserted high value acoustic resistor at its output end. A schematic set-up of the new system for measurement of the vocal tract mouth impedance is shown in Figure 3.21. The authors of [ESW97] used a high value acoustic resistor to ensure a sound velocity source close to ideal. Therefore, for determination of the impedance only the pressure signal must be recorded at the end of the horn. This is done with a miniature microphone positioned close to the *vestibulum oris* of the speaker under test.

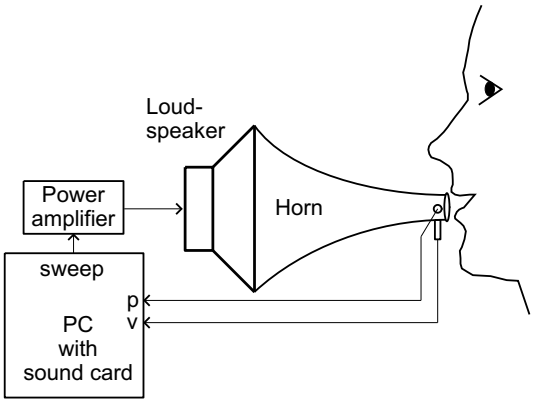


Figure 3.21: Measurement set-up

Preliminary experiences with a low output impedance using a similar set-up as reported in [ESW97] have shown that the velocity at the *vestibulum oris* was significantly altered by the load of the vocal tract, although horn and resistor should prevent such variations. Increasing the value of the acoustic resistor reduced the effect but also

the output power and, consequently, the signal-to-noise ratio. The improvement of the new setting is to measure the velocity directly by a miniature sensor (Microflown [dB02]) close to the microphone (Sennheiser KE 4 211-2, cf. Figure 3.21). This has two advantages: 1.) reduction of power required at the loudspeaker and 2.) miniaturization of the measurement set-up. The signal is processed identically for both

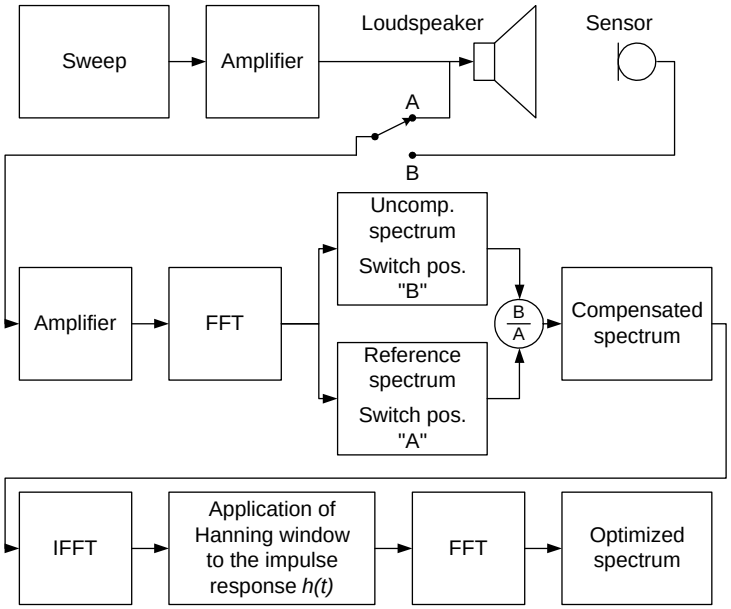


Figure 3.22: Signal flow of impedance measurements

sensors following the signal flow given in Figure 3.22. In the Figure, FFT represents the fast Fourier transform and IFFT denotes the inverse fast Fourier transform that yields the impulse responses h of the measured system. The acoustic point impedance $Z(f)$ is defined as a quotient of pressure and velocity spectra:

$$Z(f) = \frac{\mathrm{FFT}(h_p(t))}{\mathrm{FFT}(h_v(t))} \ . \tag{3.18}$$

For measurement of the impulse responses of sound pressure $h_p(t)$ and velocity $h_v(t)$, a technique using a swept sine as excitation signal is applied [MM01]. An advantage of the sweep technique is that harmonic distortions can be cancelled out by windowing the impulse response (here: Hanning window, 0..20 ms). Thus, minor changes of the mouth position or the vocal tract configuration during measurement do not reduce the overall signal-to-noise ratio since the distortion only affects the frequency range corresponding to the instant when the configuration is changed. The values of the impedances are expressed as a ratio to the unloaded, free-field impedance. The upper frequency range of the velocity sensor used allows reading of resonance frequencies,

relative amplitudes and quality of the resonances up to approx. 5 kHz. Using a PC (800 MHz), measurements with a frequency spacing of 2.7 Hz at a sampling rate of 44.1 kHz and a repetition rate below 1 Hz are possible.

The apparatus has been applied to the measurement of German long vowels and selected consonants, see Table 3.2 on page 68. In order to control the steadiness of phonemic positioning of the articulatory organs during the mutely executed VTMI measurements, ultrasonography of the tongue was carried out simultaneously with the other measurements. A 3.5/5.0 MHz probe belonging to a common 100°-sector ultrasonograph was placed in the submental region imaging the air-soft part contour of the tongue dorsum. Details of this method were published previously [SSH+81, WBH+90]. All tongue positions of the examined phonemes were visualised in the mediosagittal plane.

Measurements and Results

Resonances can be visualised by plotting the absolute value of the impedance ratio $|Z/Z_0|$ vs. frequency. All values are normalised to values measured without vocal tract attached.

In Figure 3.23 the result of a measurement on the bent aluminium model as shown in Figure 3.17 on page 61 is presented. The thin line indicates the absolute value of the spectrum of the sound pressure measured in a position equivalent to the *vestibulum oris*, the dotted line shows the absolute value of the velocity measured at the same location and the thick solid line represents the absolute value of the acoustic point impedance.

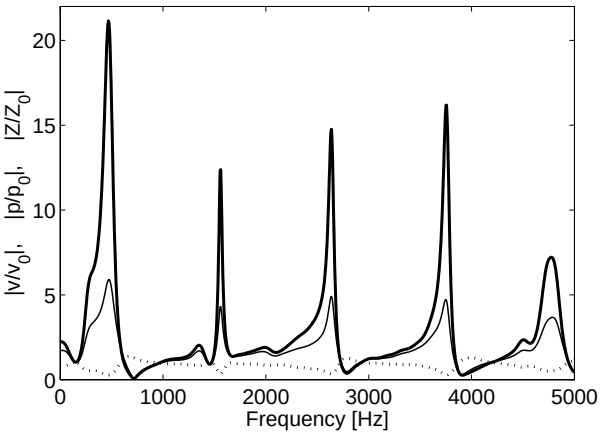


Figure 3.23: Measurement of velocity, sound pressure and impedance on a bent vocal tract model

The resonance frequencies can be identified as local maxima of the impedance curve. It can be observed that the impedance curve yields a significantly higher peak amplitude than the pressure curve. For verification of the reliability of the VTMI method, measurements using this method were compared to results from VTTF measurements on the bent aluminium model. The model has been excited with a swept sine at the glottis while simultaneous recordings of the sound pressure were performed on both ends of the cavities. In Figure 3.24 a comparison of both methods is depicted. As in the previous

experiment, the measurements were carried out on the bent aluminium model but with open inlet and outlet.

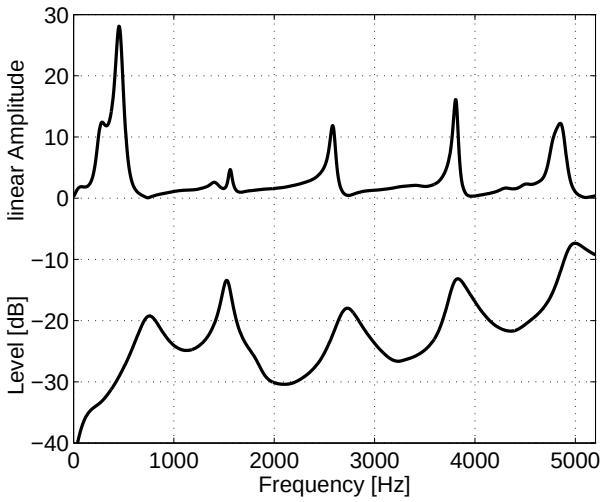


Figure 3.24: Comparison of VTMI (top) and VTTF (bottom) measurements on the bent aluminium model

The VTTF has been measured using the method for external excitation described in section 3.5.1. The upper line represents the VTMI and the lower line shows the VTTF of the same configuration. Note that the upper curve is depicted with linearly scaled ordinate, whereas the lower curve is scaled in dB. The resonance frequencies in the range between 1000 Hz and 4000 Hz match well, whereas the resonances frequencies below and above this range are lower when the VTMI method is applied. These differences might be caused by different "glottis" and radiation impedances at the openings of the cavities where the measurement head or the microphones were applied [Fla65].

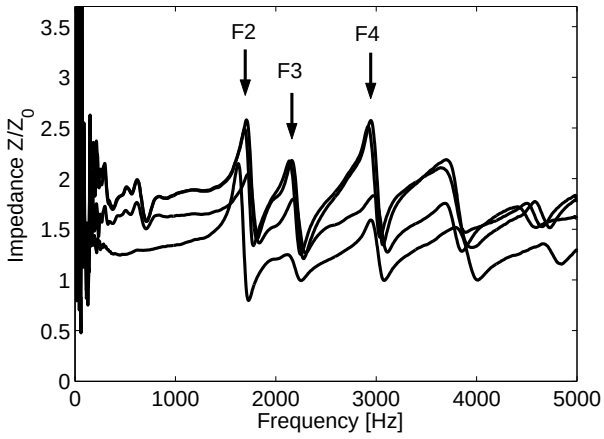


Figure 3.25: Plot of four measurements of the VTMI ratio, vowel [y:]

The reproducibility of measurements on a human subject is shown in Figure 3.25. Four subsequent measurements have been carried out while the subject articulated the vowel [y:]. The arrows indicate the measured resonance frequencies that identify the formants of the vocal tract. The resonances vary only a little between the measurements but some curves are shifted vertically with respect to others, indicating a

different distance between measurement head and mouth opening.

In Figure 3.26 a comparison between measured VTTF and VTMI for the vowel [i:] is given. The measurements have been carried out on a male subject, age 33.

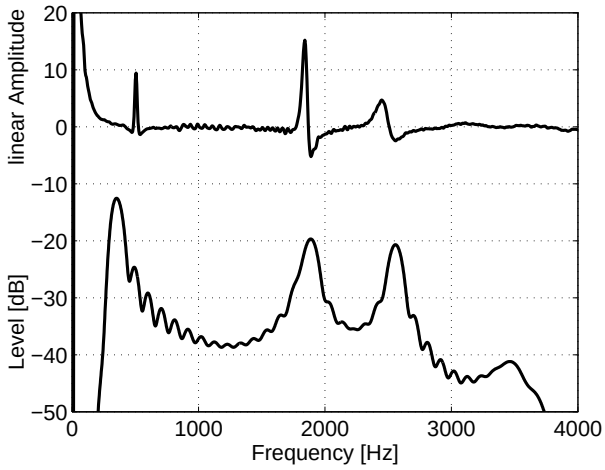


Figure 3.26: Comparison of VTMI and VTTF measurements of vowel [i:]

The signal-to-noise ratio during the VTTF measurement causes the ripple on the lower curve. As in Figure 3.24, the frequencies of the resonances measured with both methods match well.

As an example of impedance measurements of a human vocal tract with different glottis impedance conditions, in Figure 3.27 the result of three impedance measurements of a subject articulating the sound [a:] without phonation with glottis open, without phonation with glottis closed and during phonation is illustrated.

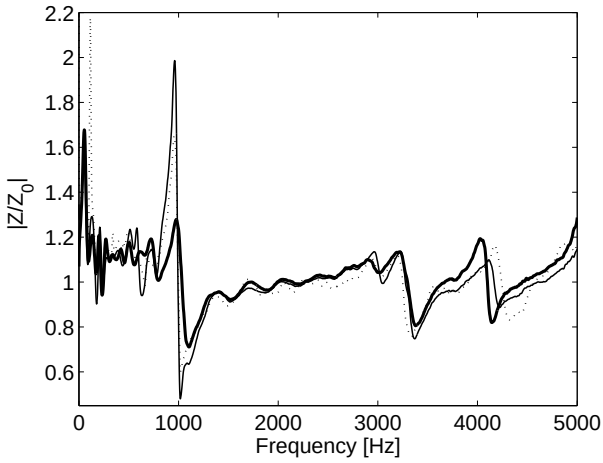


Figure 3.27: Impedance measurement with glottis open (thick solid), glottis closed (thin solid) and during phonation (dotted) , vowel [a:]

Two more kinds of measurements were performed with the same subject. In the first experiment, the pronounced phoneme was recorded four times via a microphone close to the subject's mouth. The subject was asked to hold the pronunciation of the expressions on the bold speech sounds listed in Table 3.2.

Table 3.2: Mean values and standard deviations (in brackets) of formant frequencies of the vowels (F2 & F3/F4*) and poles in the spectra of the consonants, obtained by four independent measurements using LPC and VTMI methods.

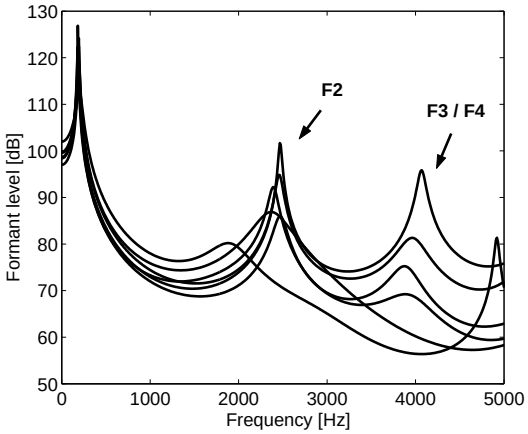
Speech sound Expression	LPC [Hz]	VTMI [Hz]	LPC [Hz]	VTMI [Hz]
Vowels	2^{nd} Formant		3^{rd} or 4^{th} (*) Formant	
[a:] **a**ber	781 (60)	1110 (15)	3176 (238)	3515 (85)
[ɛ:] **Ä**hre	1680 (86)	1520 (20)	3335 (157)*	3450 (20)*
[e:] b**e**ten	2062 (51)	2450 (50)	3251 (42)	3250 (50)
[i:] M**ie**te	2261 (118)	2600 (50)	3187 (65)	3275 (25)
[y:] H**ü**te	1901 (67)	2225 (25)	3036 (126)*	3015 (35)*
Consonants	1^{st} Pole		2^{nd} Pole	
[ʃ] Sa**sch**a	2266 (67)	2125 (75)	4816 (134)	4750 (50)
[x] a**ch**	1076 (72)	1045 (25)	3742 (38)	3450 (250)
[j] ja**j**a	3300 (60)	3300 (100)	4482 (235)	4150 (50)

The recording of the uttered sounds was evaluated by a linear predictive coding (LPC) analysis using the COLEA[2] package for the program language MATLAB. In the second measurement, the VTMI was measured four times under sonographic observation to ensure the articulatory configuration did not change between the measurements. Still, the invariability of the vocal tract configuration could only be verified in the imaged plane. As a consequence of the small size of the loudspeaker used in the measurement set-up, only a determination of the resonances above 500 Hz could be achieved using the VTMI method. However, the use of a larger loudspeaker should allow a derivation of the first formant (cf. [ESW97]). The results of the measurements are summarised in Table 3.2.

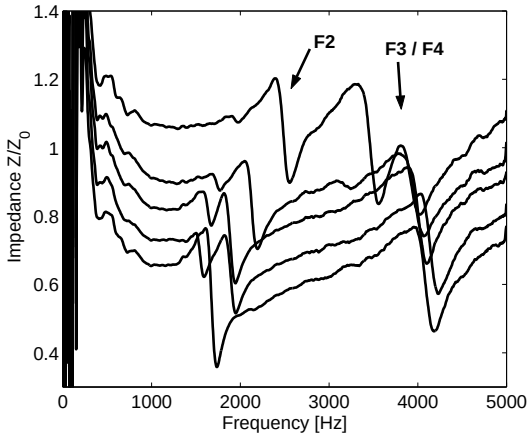
The values indicate a satisfactory reproducibility for each method with exception of a few phonemes. A comparison of LPC and VTMI analysis should result in similar – though not necessarily identical – values for the formants or poles. Nevertheless, in some cases significant differences have been observed.

Another example of a comparison of the two methods is given in Figure 3.28. The transition of formants of the vowel [i:] pronounced in several tongue positions changing successively from ventral to dorsal (here: front to back) are presented as LPC analysis (a) and VTMI measurement (b). The upper curves represent the spectra of the vowel [i:] (ventral tongue position) whereas the lower curves point out a shift of F2 towards lower frequencies caused by a dorsal position of the tongue. The upper

[2]available on the internet at `http://www.eeng.dcu.ie/~speech5/matspeech.html`



(a) LPC analysis (order 8, 50 ms window length)



(b) VTMI analysis, curves successively shifted by 0.1

Figure 3.28: Formant (a) and impedance (b) curves for a sequence of [i:] with tongue moved from ventral to dorsal position

formants F3 and F4 are difficult to track using LCP analysis but remain visible using the VTMI method.

Conclusions

It was demonstrated that the VTMI method is able to visualise functional vocal tract characteristics during articulation. The method presented yields fast and reliable impedance measurements in the frequency range from $\sim 500\,\text{Hz}$ to $5\,\text{kHz}$, even with a single sample of duration 166 ms. However, results from LPC analysis and VTMI measurements seem to differ significantly for some phonemes. This may be caused by undetected tongue movements not visualised in the selected sonographic imaging plane or by differences of the glottis impedance with or without phonation. Furthermore, a change of configuration between recording and VTMI measurement cannot be excluded with certainty. The problem should be solved by using an optimised set-up allowing an on-line voice recording and VTMI measurement in very fast sequence or even simultaneously, as described in [ESW97].

Further studies are intended to work out different clinical applications of the method. The preliminary results shown for the vowel [i:] indicate that the method is able to differentiate among various kinds of vocal tract configurations. Therefore additional studies are intended to examine the impedance characteristics of the vocal tract in groups of patients suffering from hyperfunctional dysphonia compared to groups of healthy adults. The voice production of these patients is often characterised by an impaired vocal tract function caused by the pathological use of the

false vocal chords or by a functional backward dislocation of the tongue. In addition, examinations in healthy children are proposed to be useful to obtain clinical data for a prospective application of the method in the articulatory training of deaf children. In this context, it is of great advantage that this method is able to visualise the vocal tract function even in mutism [DSW96, DSW97]. The dynamic derivation of the vocal tract transfer function or the vocal tract area functions from the vocal tract impedance at the mouth during speech might be a future task.

3.6 Discussion

The vocal tract has few but quite complex properties that are difficult to model. The proposed models are suitable for calculation of the vocal tract transfer functions, both as separate modules or within a waveguide. The measurements are difficult to compare to the calculations since the MRI data has been obtained from individuals who are not available for the invasive measurements. However, good agreement could be found for the resonance frequencies for all vowels. The damping is a critical value that needs more detailed investigations. For the CTIM cone model the damping is a parameter that directly determines the stability of the calculations. A drawback of the algorithm is its unstable behaviour for segments with increasing diameter. This problem is known [MAC88, Smi96] and can be avoided for some vowels by choosing an adopted set of parameters for the segments. An improved algorithm should solve the problem of instability for critical VT configurations.

Chapter 4

Noise generation

In this chapter the relevance of noise generation for voice synthesis is discussed. In section 1.1.3 it has been shown that an analysis of a voice signal always reveals a significant noise content. The vocal fold signal generation using the models described in section 2.2 do not include noise generation and therefore sound unnatural. During voice production noise is generated either when an air stream is directed against an obstacle or when a divergent configuration or a discontinuity of the bounding walls causes a laminar flow to separate from the boundary. In both cases the laminar flow is disturbed and vortices are shed that travel some distance and then break down and dissipate their energy in turbulent air movement.

For a natural sounding voice synthesis it is necessary to take into account more aspects of the aerodynamics of the voice generation process than just Bernoulli's law. Noise generation during voice production may occur whenever a non-laminar stream of air propagates. This may happen in two cases:

- constrictions within the vocal tract
- jet formation at the glottis.

The perceptual relevance of the first case is evident because the oscillator-filter model cannot produce unvoiced sounds such as consonants or whispered voice. The second case has been investigated by D. Hermes [Her91], P. R. Cook [Coo90] and recently by P. J. B. Jackson [Jac00, JS00] and D. J. Sinder [Sin99]. From Lighthill's analogy it can be derived that sound generation by turbulences is most efficient near changes in geometry [Hir92]. In this thesis only aspiration noise will be considered.

4.1 Noise sources

From a perceptive point of view, two different noise types can be distinguished: noise that is generated in voiceless sounds such as consonants or noise that is part of a

voiced sound such as vowels. For singing voice synthesis the consonants are mainly used to connect vowels and play a minor role for the timbre of a voice, whereas for speech synthesis the consonants carry most of the information.

4.1.1 Fricatives

Noise sounds generated in the vocal tract or at the mouth are called fricatives (from Latin *fringere*: to break). This sound occurs when flow passes through a narrow passage or hits an obstacle. To get an impression of the sound you can form a jet with your lips and direct it against a close object. Examples for such sounds are the phonemes /ʃ/ (**sh**ake), /f/ (**f**eel) or /s/ (**s**ee). Fricatives are very important for speech synthesis and analysis. Since the comprehensive work of Shadle [Sha85] several attempts have been made to measure [BMB+96], characterise [Jac00] and model [Sin99] fricatives. For singing voice synthesis the use of fricatives is restricted to transient sounds, combined voiced and fricative sounds, or non-voiced sounds. In this work fricatives will not be considered.

According to M. Liu and A. Lacroix [LL98] the generation of fricatives can be modelled by the KL model. The insertion of secondary sound sources, like fricative noise, between the glottis and mouth segments adds two poles to the frequency response of the VT.

4.1.2 Aspiration noise

Even in voiced sounds a significant part of the signal consists of noise. The origin of this noise cannot be explained by the sound generation process that has been described in chapter 2.

Its origin is due to vortex shedding at the vocal folds that causes turbulences downstream of the glottis. In contrast to fricative noise, aspiration noise is generated in pulses that are synchonised with the glottal cycle. In Figure 4.1 the waveform of a synthetic vowel [a:] is shown. The thick line indicates the sound pressure as produced by the vocal fold movement. The thin line is the sound pressure of the noise.

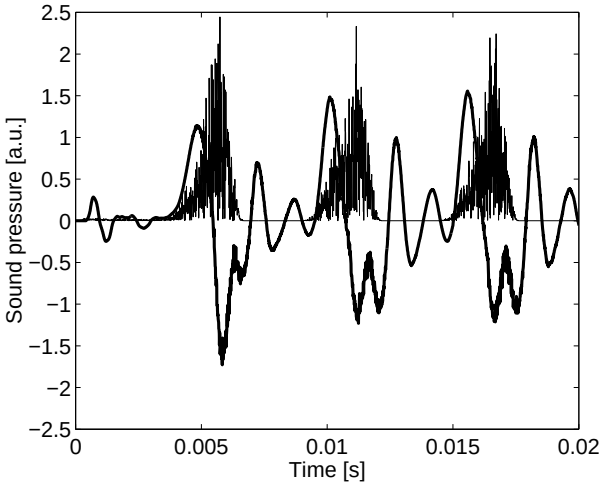


Figure 4.1: Calculated waveform and noise burst for vowel [a:]

Listening tests by D. Hermes [Her91] revealed the perceptual relevance of relative amplitude and relative phase of noise pulses and glottal pulses. An integration of noise and harmonics to a unique sound only occurs when

- the noise is modulated by the glottal pulse,
- the energy within both components is of same order, and
- the phase difference does not exceed $T/8$.

Modeling of the pulsed noise has been investigated earlier. However, a model that generates aspiration noise using a physically motivated algorithm has not been implemented for a singing voice model.

4.2 Noise models

For the generation of noise various models have been reported in literature.

J. Liljencrants [Lil85] uses the following equation for the generation of a momentary noise flow u_n

$$u_n = U_0 \left(1 - \frac{1}{x}\right) c_n K_n \tag{4.1}$$

with the quotient x that relates the actual Reynolds number Re to a critical Reynolds number Re_c:

$$x = \left(\frac{Re}{Re_c}\right)^2 = \frac{4}{\pi}\left(\frac{\rho}{\mu Re_c}\right)^2 \frac{U_0^2}{A} \; . \tag{4.2}$$

At Reynolds numbers below Re_c no noise is generated. A value of $Re_c \approx 1800$ has been determined by experiments on plastic models. In equation (4.1) the random number c_n with $c_n \in \{-1..1\}$ models the amplitude fluctuations of the noise signal, and K_n is a factor describing the efficiency of the energy transfer from the average (DC) flow U_0 into the noise flow u_n. For K_n an empirically determined value of 0.01 is assumed.

P.R. Cook modelled pulsed noise generation by implementation of an aerodynamically motivated noise generation module into the signal chain of a complex oscillator-filter model [Coo90]. This algorithm takes into account the pulsed noise and relates it to flow parameters (U, v). However, the assumption of free field boundary conditions that yield a radiated power proportional to U^8 does not apply for the duct-like environment of the glottis [Hir92].

D.J. Sinder recently developed a noise model [Sin99] based on simple equations that were obtained from fluid dynamic theory. The main idea is the description of vortices by a small number of parameters that can be derived from previously calculated measures, for example, flow velocity or geometrical boundary conditions.

Equation (4.3) gives the relation between the expectation value for the shedding interval $E\{T_{shed}\}$ and vortex diameter D, Strouhal number St and jet velocity U_j.

$$E\{T_{shed}\} = \frac{D}{StU_j}. \tag{4.3}$$

The Strouhal number $St = \frac{f_{shed}D}{U_j}$ is a dimensionless measure for the ratio of acceleration due to the unsteadiness of the flow and convective acceleration due to the non-uniformity of the flow [Hir92]. It corresponds to the mean frequency of the generated noise spectrum.

The vortex production rate increases when the flow increases or when the diameter of the opening decreases.

$$P_{noise} = \frac{-\pi d\rho_0}{A}[\vec{\Gamma} \times \vec{v} \cdot \vec{U}]. \tag{4.4}$$

In equation (4.4) $\Gamma = \frac{1}{2}v^2T$ is the vortex rotation and $\vec{U}$ denotes the normal component of the surrounding flow.

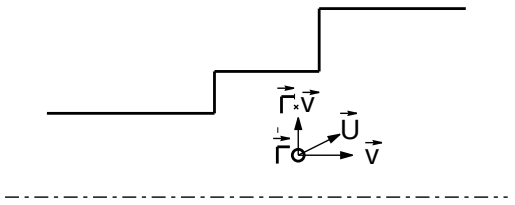


Figure 4.2: Relation between flow, vorticity and vortex velocity

The relations between the geometry in the vocal tract and the flow properties are illustrated in Figure 4.2. The sound pressure is proportional to the sine of the angle between the flow direction U and the jet velocity v. An interesting aspect of this approach is the dependence of the noise generation on the geometry of the first supraglottal segments. For sharp discontinuities in this area a strong noise contribution is expected. In 4.4.3 this prediction is investigated experimentally.

4.3 Implemented model

The noise generation algorithm used in this work is based on the work of D. J. Sinder [Sin99]. Two significant differences to Sinder's algorithm have been implemented. At first, the assumption of energy loss of the vortex during propagation has been added. It is assumed that the vortex energy is reduced exponentionally with growing distance at an estimated rate of 40 dB for a travel distance of $4 \cdot D$. As a second difference the source smoothing algorithm has not been implemented.

In this implementation each vortex is programmed as a struct in which the parameter values are the struct elements. The algorithm for vortex generation is shown in Figure 4.3.

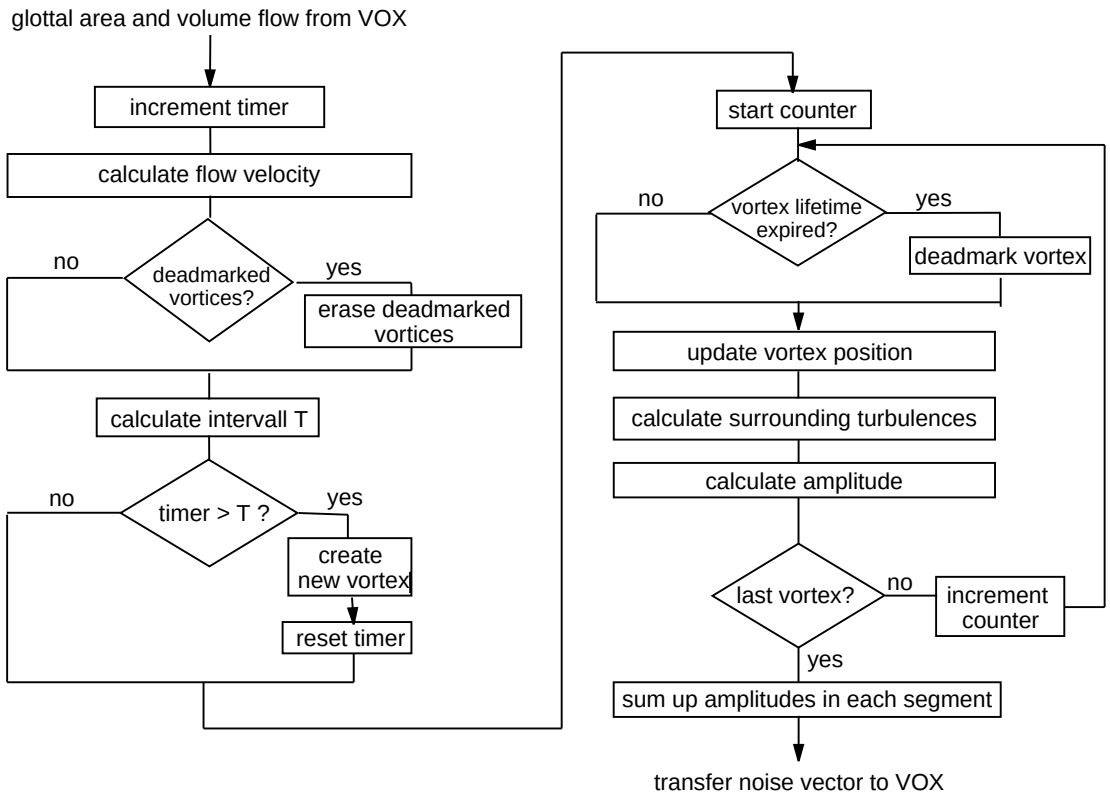


Figure 4.3: Signal flow of the noise generation module

From the actual parameters flow, EAF and glottal minimum area the flow velocity is calculated [Hir92]:

$$v = \frac{U_j}{A_{min} \cdot 1.1} \ . \tag{4.5}$$

The shedding interval T_{shed} is calculated according to equation (4.3) and, if the time difference between the last vortex generation exceeds T_{shed}, a new vortex with initial velocity 0 $\frac{m}{s}$ is generated. For each vortex, position and diameter are then calculated from the actual flow parameters. According to equation (4.4) the contribution of the vortex to the turbulent sound is calculated. During propagation the object properties are updated until the conditions for the life of the vortex are expired. A detailed description of the implementation of the model can be found in [Sel01].

Figure 4.4 exemplarily shows the propagation of generated vortices in the vocal tract. The vortices are represented as lines of variable length and intensity. The diameter of the vortices equals the length of the lines whereas their intensity is indicated by gray-scaled colours. The darker the lines, the more noise is produced. The Figure illustrates that a sharp discontinuity in geometry causes an increased sound production, as expected from equation (4.4).

The *glottis* is represented in Figure 4.4 as the area below 0 on the abscissa. During calculation of voiced sounds, the minimum area change and the noise production can be monitored. The algorithm is integrated into the combined voice model described in chapter 6.1.

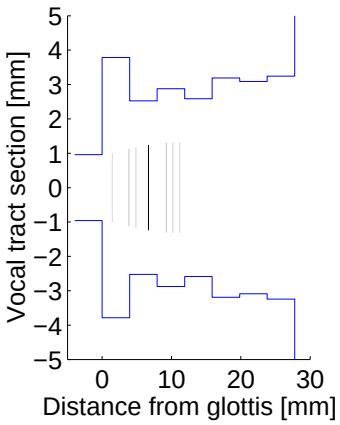


Figure 4.4: Vortex shedding in the vocal tract

4.4 Measurements

The analysis of the noise contents in a voice signal is a difficult task. The most difficult problem is the time-varying fundamental frequency that does not allow to just calculate a long time FFT and then consider the harmonic and non-harmonic part of the signal. Different methods are available for the analysis the non-periodic part of voice, e. g. reviewed in [Jac00] or [Coo90].

4.4.1 Time-domain analysis

The noise analysis can be performed in the frequency domain or in the time domain. The glottal-to-noise excitation ratio (GNE) is a measure that has been proven to yield reliable results for the characterization of the noise content in a voice signal [MGS97]. The signal is analysed using a correlation method that has been implemented for the characterization of hoarseness by determination of the two measures irregularity and GNE. A graphic representation of these parameters is the Göttingen hoarseness diagram that can be used for documentation of the course of a voice therapy[1].

4.4.2 Frequency-domain analysis

The method used in this work for the frequency-domain analysis of voiced sounds is an implementation of the PSHF algorithm by Jackson [JS98]. Among the frequency-domain methods, the pitch-scaled harmonic filter (PSHF) is a recently published method [Jac00] that requires the measurement of the period lengths for the separation of harmonics and noise. Therefore, it can be applied to voiced sounds only.

Prior to the application of the PSHF algorithm, the exact length of each period has to be determined. This is achieved by application of a bandpass filter to the voice signal around the estimated fundamental frequency and subsequent indexing of the zero-crossings. Then the PSHF method applies overlapped FFT over four periods, subsequently shifted by one period length. The harmonic part is present in every 4^{th}

[1] Information can be found on the internet at
`http://www.physik.uni-oldenburg.de/Docs/medi/hoerz/`

bin of the spectrum whereas the non-harmonic part resides in the three other bins. The algorithm allows the separation of the harmonic and the non-harmonic part of the signal.

An advantage of this method over other methods is the high immunity against pitch changes that are characteristic for the human voice [JS98]. However, if the HNR is very low the method fails because the period lengths cannot be determined anymore. In this case, the application of cross-correlation methods is appropriate [MGS97]. A comparison of both methods is given in the following section.

4.4.3 Results

To analyse the noise content in the voice, recordings of 13 healthy subjects were performed in an anechoic chamber using high quality measurement equipment. In Figure 4.5 the results of a PSHF analysis of 11 non musically trained subjects (indices 1..11) and 2 experienced singers (indices 12 and 13) pronouncing a number of vowels are presented. On the abscissa the index of the singer is given, and to the right mean values

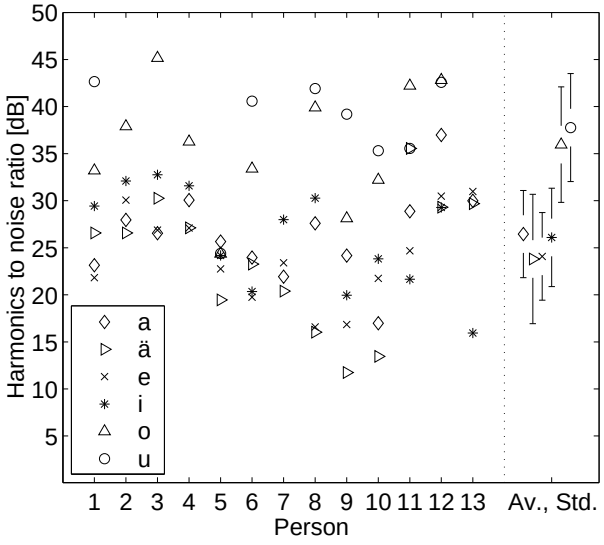
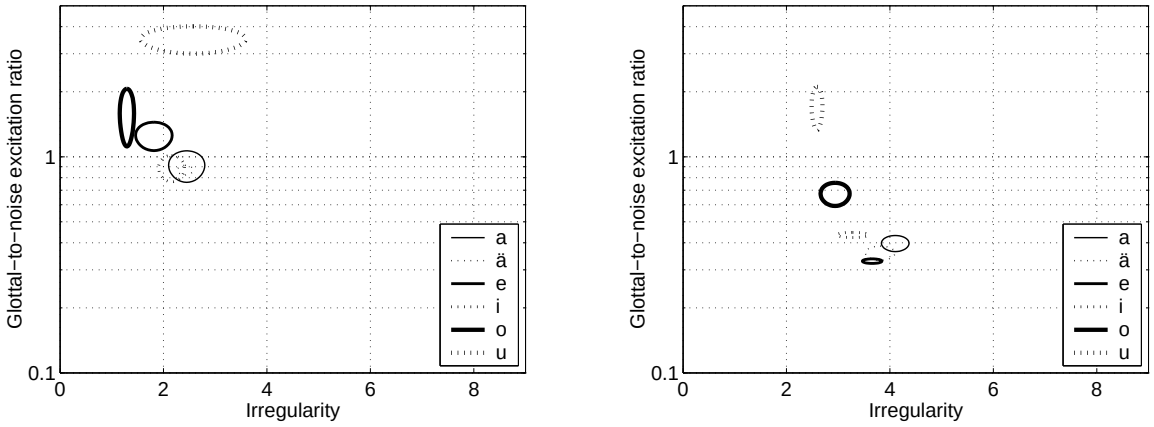


Figure 4.5: Results from a HNR-Analysis using PSHF

with standard deviations for the different vowels are presented. Measurements that could not be evaluated due to disturbant noise (e. g. direct breathing into the microphone) are excluded. From the mean values it can be concluded that the HNR value differs significantly for different vowels, the vowels [o] and [u] are well separated from the other sounds. This analysis gave reason for a more detailed investigation of the correlation between the sung vowel and the noise content. The aim of the described measurements is the verification of the implemented noise generation algorithm.

For an impression of the difference in noise content between a musically trained and a non musically trained singer, in Figure 4.6 the results of a hoarseness diagram

analysis of an experienced singer (a) and an unexperienced singer (b) are presented. It



(a) Experienced singer (b) Unexperienced singer

Figure 4.6: Hoarseness diagrams of an experienced (a) and a less experienced singer (b)

can be seen that the cluster of curves is located towards less irregularity and towards less noise contents for the experienced singer. The vowels [o:] and [u:] contain less noise content for both singers.

For verification of the results that were obtained with the PSHF algorithm, a second analysis of the same voice signals has been performed using the GNE algorithm [MGS97]. In Figure 4.7 the results of GNE analysis of the same voice signals are presented. On the abscissa the index of the singer is given, and to the right mean

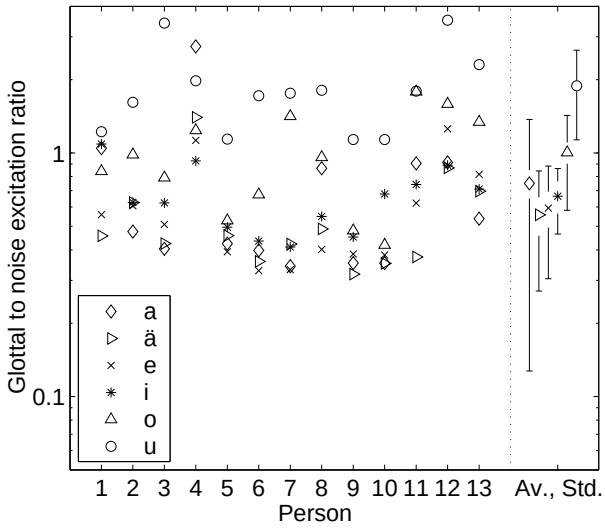
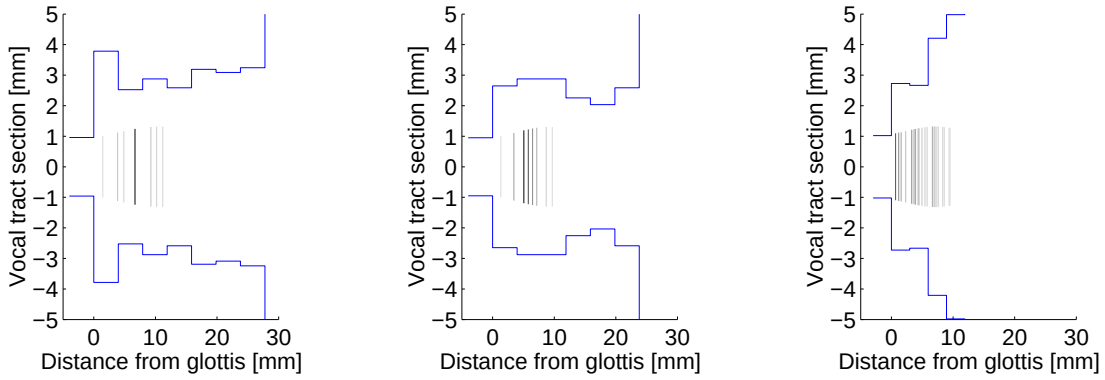


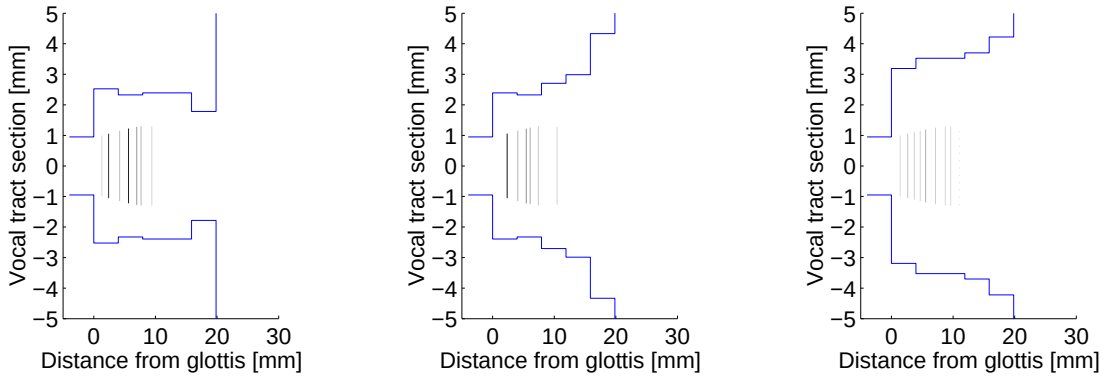
Figure 4.7: Results from a GNE analysis using the Göttingen hoarseness algorithm

values (Av.) with standard deviations (Std.) for the different vowels are presented. Although the GNE algorithm works on a different way from the PSHF algorithm, the data presented on Figures 4.5 and 4.7 show similar trends. The average values

for the noise content are grouped in two domains: for the vowels [a] and [u] a significantly lower noise content can be observed. In Figure 4.8 the supraglottal vocal tract area is shown for different configurations. The vocal tract area data is taken



(a) vowel [a:] (b) vowel [æ:] (c) vowel [e:]



(d) vowel [i:] (e) vowel [o:] (f) vowel [u:]

Figure 4.8: Supraglottal equivalent area functions of the vocal tract for different vowels

from [ST96] with exception of the data shown in Figure 4.8 (c) which has been taken from [KWMP00]. All plots are snapshots that were taken during a simulation with the standard configuration (cf. section 2.3.4) when the equivalent glottis diameter (area segment left from zero on the abscissa) was 2 mm during the closure phase. The equivalent area data is rather smooth and slightly divergent for the vowels [o] and [u] but exhibits rather strong discontinuities for the vowels [a],[æ] and [e].

4.5 Discussion

The generation of aspiration noise is important for the naturalness of voice.

The noise content in measured and simulated voice signals varies with the vocal tract configuration which is in line with the vortex theory. The noise content in front of the mouth also dependents on the mouth area, because the frequency dependent radiation impedance is bigger for larger mouth openings. The mouth area is rather big for the vowels [a:], [æ:] and [e:] but small for the vowels [o:] and [u:]. Since the mouth opening for the articulation of the vowel [a:] is biggest, one would expect the maximum noise content. However, the measurement results indicate that the noise content is even stronger for the vowels [æ:], [e:] and [i:].

The use of the implemented noise model is not restricted to the generation of aspiration noise at the vocal folds. With minor modifications, the same algorithm could be applied to noise generation at arbitrary positions within the voice organ. This would allow simulation of fricatives like [ʃ], [s] or [x].

Chapter 5

Radiation

The sound wave that has travelled from the glottis to the mouth, is partly reflected at the mouth opening and partly transmitted into the surrounding air. The reflected part contributes to the superposition of waves propagating in the vocal tract. The importance of this effect has been discussed in section 3.2.2. The transmitted part is responsible for the radiation and perception of the generated sound. Both, reflected and transmitted sound waves are affected by the discontinuity between vocal tract and an assumed free field around the human body. The frequency-dependent effect is characterised by the radiation impedance that is described in section 5.1.

Directivity is defined as the spatial distribution of sound energy around a sound source. Whereas for rather simple models like a baffled tube that radiates in free field, the sound radiation is known from measurements or calculations (see, for example, [Sch66]), a detailed investigation of the radiation characteristics of the singing voice is still lacking. Measurements of the radiation characteristics have been carried out by J.L. Flanagan [Fla60] as well as by J. Meyer and H. Marshall [MM84] who demonstrate the general radiation behaviour of an artificial or a human singer. In section 5.2 new measurements of the singer's directivity are presented.

5.1 Calculation

Several models exist for the description of the radiation impedance of the singer. One commonly used approach is the description of the actual radiation by assuming an oscillating piston in an infinitely large baffle. This model can be further simplified as an oscillating sphere with a surface that equals the piston area [Sch99].

The characteristic impedance of such a sphere is given by

$$Z(r) = \rho c \frac{jkr}{1 + jkr} \ . \tag{5.1}$$

The transformation of equation 5.1 in the time domain [Kro95] yields the exponential

reflection function

$$R_m = -\frac{c_0}{2r} \cdot e^{-\frac{c_0}{2r}t}\varepsilon(t) \tag{5.2}$$

with the Heaviside unit step function ε. For the case of a circular piston in an infinite wall the radiation impedance can be calculated as follows [Fla65]:

$$Z = 1 - 2\frac{J_1(kd)}{kd} + 2j\frac{K_1(kd)}{(kd)^2} \ . \tag{5.3}$$

J_1 denotes a Bessel function of 1^{st} order, K_1 is a related Bessel function, $k = \frac{2\pi f}{c}$ is the wave number, and d is the diameter of a circle that is equivalent to the mouth area. A plot of the magnitude (Abs), the imaginary (Im) and the real (Re) part of the

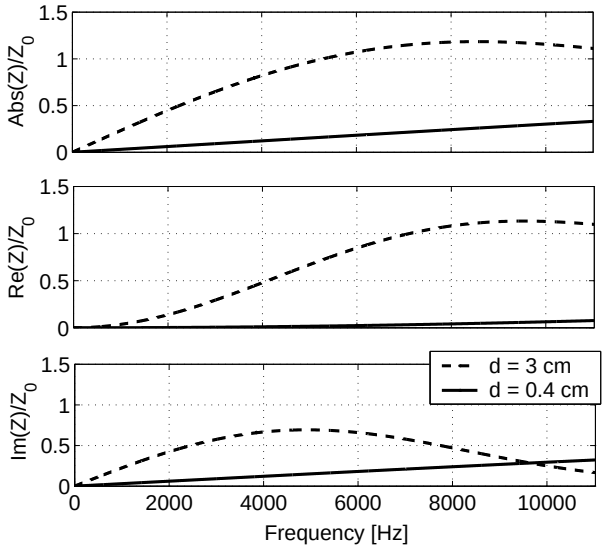


Figure 5.1: Radiation impedance of a baffled piston

radiation impedance normalised to the free field impedance Z_0 for two typical mouth opening areas is presented in Figure 5.1. The solid line corresponds to a closed mouth as for [u:], whereas the dashed line represents a widely opened mouth as for [a:].

For big openings the magnitude of the radiation impedance is not linearly increasing with frequency but shows convergence for frequencies above 5000 Hz towards 1. In most radiation models the frequency dependence is described using the simple equivalent electric circuit shown in Figure 5.2. The values for the resistance R and the inertance L were determined by Flanagan [Fla65] to be

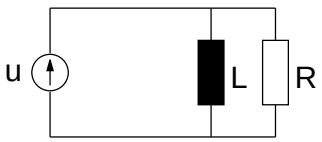


Figure 5.2: Electric circuit for mouth radiation

$$L = \frac{8\rho}{3\pi^2 r} \text{ and } R = \frac{128\rho c}{9\pi^3 r^2} \tag{5.4}$$

where c is the speed of sound, ρ is the air density, and r is the lip radius. A significant deviation from the actual values is expected for frequencies above 3 kHz.

5.2 Directivity measurements

The directivity measurement of sound sources is a well-known method in the fields of electroacoustics and musical acoustics. The measurement aims at describing the sound radiation characteristics of the sound source in terms of a graphical or analytical description of the distribution of emitted sound power. Parameters of the evaluation are elevation and azimuth angles and the frequency band.

The directivity measurement of mechanical musical instruments yields high accuracy due to the repeatable broadband excitation either with an electric [MM95] or a mechanical signal. Measurements of the directivity of a singer cannot be done in a similar way because of two main problems:

1. The excitation signal is neither constant in amplitude nor in pitch. Therefore, a measurement with a good reproducibility is very difficult and exhausting for the singer under test.

2. A scan of equidistant positions around the singer cannot be easily done by turning the sound source in arbitrary positions.

A method for directivity measurements of human singers is presented in the following section. For reproduction of the calculated signals and more repeatable directivity measurements an artificial singer is described in section 5.2.2.

5.2.1 Human singer

A measurement method has been developed using the voice as excitation signal [KJ99, Jer98]. The method requires two microphones M_1 and M_2, a PC with data acquisition and signal processing facilities, a turntable in an anechoic room, and a curved beam with a sledge to allow variable vertical positioning of the microphone M_1 equidistant to the singer's head. The setup for the directivity measurements is shown in Figure 5.3.

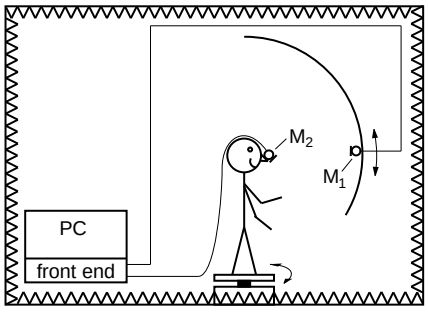


Figure 5.3: Setup for measurement of a singer's directivity

Compensation for microphone position: Due to the microphone's close position to the mouth and the head, an equalization of the stationary spectrum recorded by M_2 has to be performed. For this purpose, the artificial singer (described in the following section) was equipped with a "nose microphone" and was fed a "pre-whitened" signal that yielded a flat spectrum at a distance of two

meters. The set-up for the measurement of this compensation spectrum is shown in Figure 5.4 (left). Moreover, to the right the spectrum of the signal from the nose microphone recording is plotted.

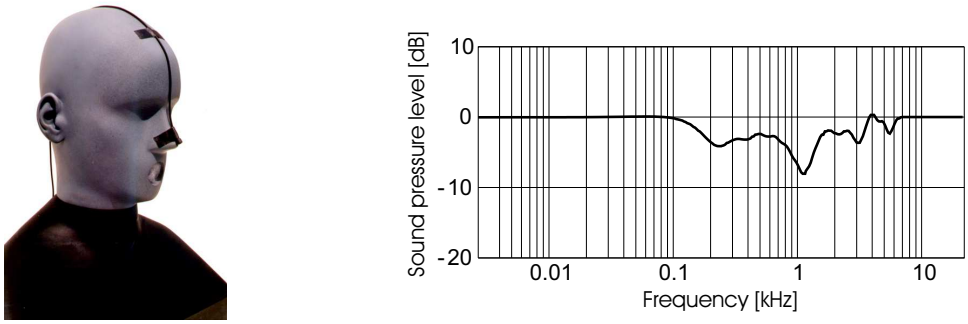


Figure 5.4: Set-up (left) and spectrum (right) for correction of the microphone position

The correction for the level loss was applied during internal processing of the signals within the measurement program.

5.2.2 The artificial singer

For the reproduction of a voice signal the radiation characteristic of the human torso plays an important role. An artificial singer has been built to verify the measurements of human singers.

The purpose of constructing an artificial singer is to achieve natural reproduction of the human voice with regard to both the frequency behaviour and the radiation properties. Preliminary investigations on a simple cylinder-shaped artificial speaker revealed insufficient sound power due to the limitations of the small loudspeaker as well as a radiation characteristic that lacks the details of a real singer's directivity. A human-like artificial singer was constructed that meets the requirements for the reproduction of a loud human voice. A photo of the singer is illustrated in Figure 5.5 (a).

The following list contains the implemented features of the artificial singer:

Technical properties

- Shape equal to the ITA dummy head [Sch95] for natural diffraction
- Two low-range speakers (Fane Studio 5M) with a 4th-order symmetric bandpass system for the reproduction of low frequencies, opening of the bass reflex tube in the neck
- One midrange loudspeaker (TPC 80 RW/4h) for radiation at the mouth

- Guided airflow from loudspeaker to mouth to avoid standing waves
- Transition frequency between midrange and bass system: 150 Hz
- Equalization of the frequency response for both channels separately with the digital controller HUGO [Kle96, SK00]

In Figure 5.5 the artificial singer is shown. The head is made of polyester by using

(a) front view (b) opened

Figure 5.5: Photos of the artificial singer

a gypsum form. The upper part of the torso (a) consists of seven handcrafted pieces of wood. The two bass speakers can be seen (b), as well as the bass reflex tube. A technical drawing of the torso is given in appendix C.

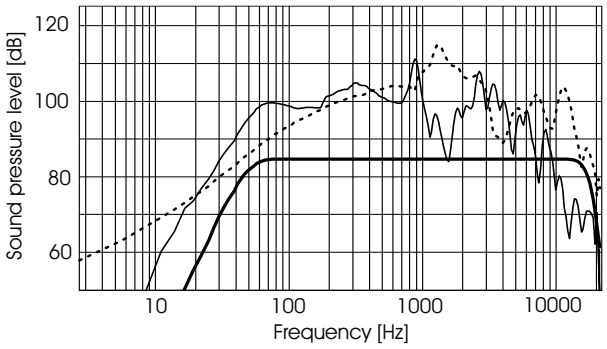


Figure 5.6: Spectrum of the artificial singer measured at 1 m distance from the mouth

The installed loudspeakers have different frequency ranges for optimum operation. In Figure 5.6 the frequency responses of the installed midrange (dotted) and the bass loudspeaker (thin solid) are depicted as well as the curve for the equalised singer (thick solid curve). The high and constant level that can be achieved with the equalised system extends 80 dB from below 70 Hz and allows reproduction of soprano to bass singing voices. Subjective listening tests confirmed the realism of the head's voice both for reproduction of voice recordings and for convoluted signals in choir auralisations using measured impulse responses [Jer98].

A sequence of directivity measurements of the artificial singer in the transverse plane (elevation = 0°) and the sagittal plane (azimuth = 0°) is shown in Figure 5.7. Due to the set-up, the radiation could be measured in the upper hemisphere only. The measurements are normalised to the 0°/0° direction. The radiation characteristics

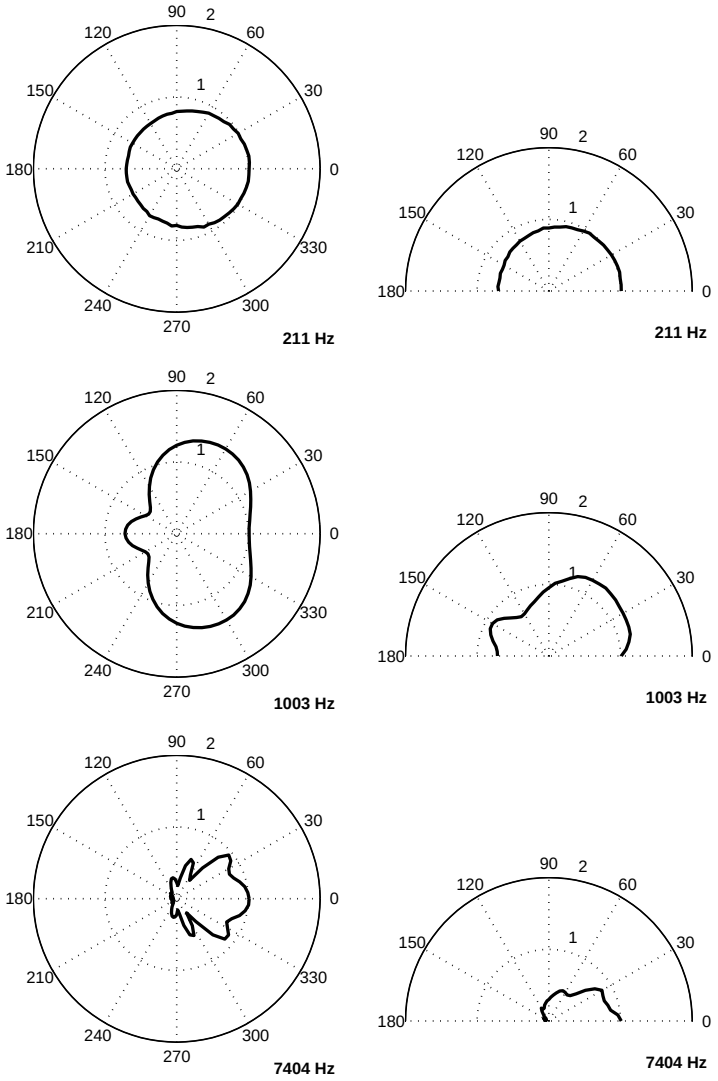


Figure 5.7: Polar plots from directivity measurements of the artificial singer

are comparable to the directivity patterns of loudspeakers. At low frequencies the radiation is almost omnidirectional. With rising frequency, sidelobes can be observed that move from front to back with radiation at the back continuously decreasing. At high frequencies (>6 kHz) the main radiation is directed to the front.

These observations are in agreement with the findings of J.L. Flanagan [Fla60], and J. Meyer and H. Marshall [MM84].

Comparison human singer – artificial singer

The Figures 5.8 and 5.9 present a comparison of the directivity between a female singer, male singer and the artificial singer in the transverse plane (left) and the sagittal plane (right).[1] The plotted radiation patterns were chosen by their similarity in order to compare the frequencies at which they occur. The plots of Fig. 5.8 indicate

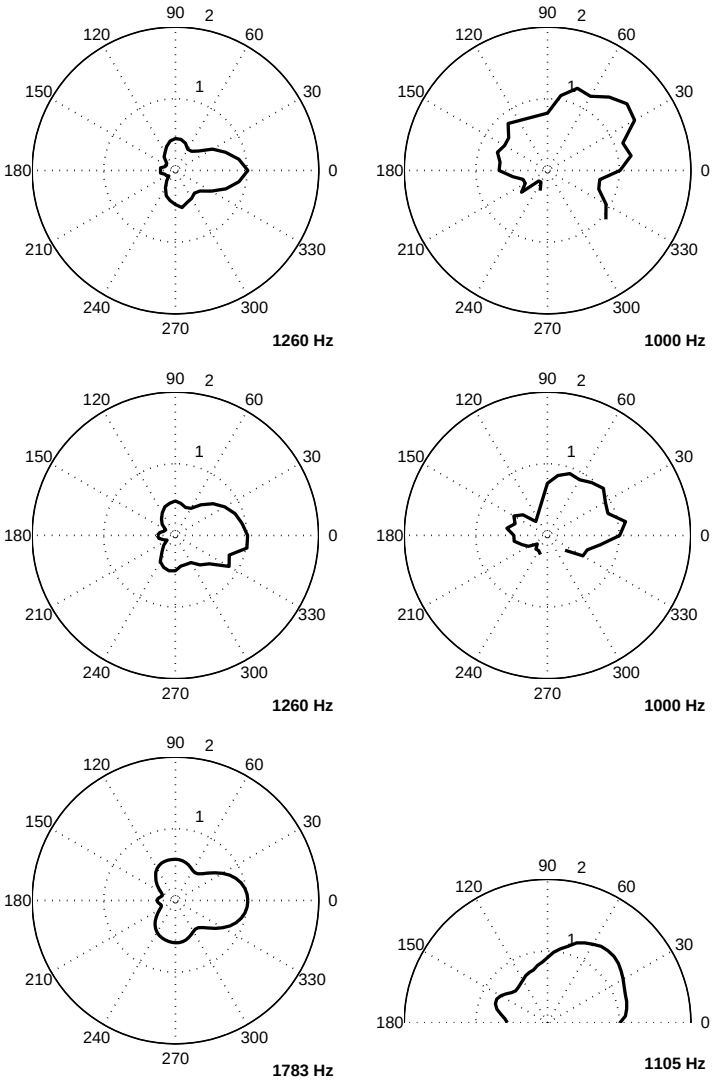


Figure 5.8: Comparison of directivity patterns of a female singer (top), a male singer (middle) and the artificial singer (bottom) around 1000 Hz

that similar radiation characteristics for the female, the male and the artificial singer can be found at different frequencies. For male and female singers, the radiation

[1]Further results (static and animated polar plots) can be found on the web page `http://www.akustik.rwth-aachen.de/~malte/directivity`.

patterns are relatively identical at the same frequency, although they occur at lower frequencies compared to the equivalent patterns of the artificial singer.

This effect can also be observed at a higher frequency range as presented in Figure 5.9. As in the previous analysis, the directivity patterns of the human singers occur at lower frequencies compared to those of the artificial. In addition, at higher frequencies differences between male and female singers are found. In the sagittal plane, the described pattern cam be seen at a lower frequency for the male singer compared to both the female and the artificial singer. Another difference was ob-

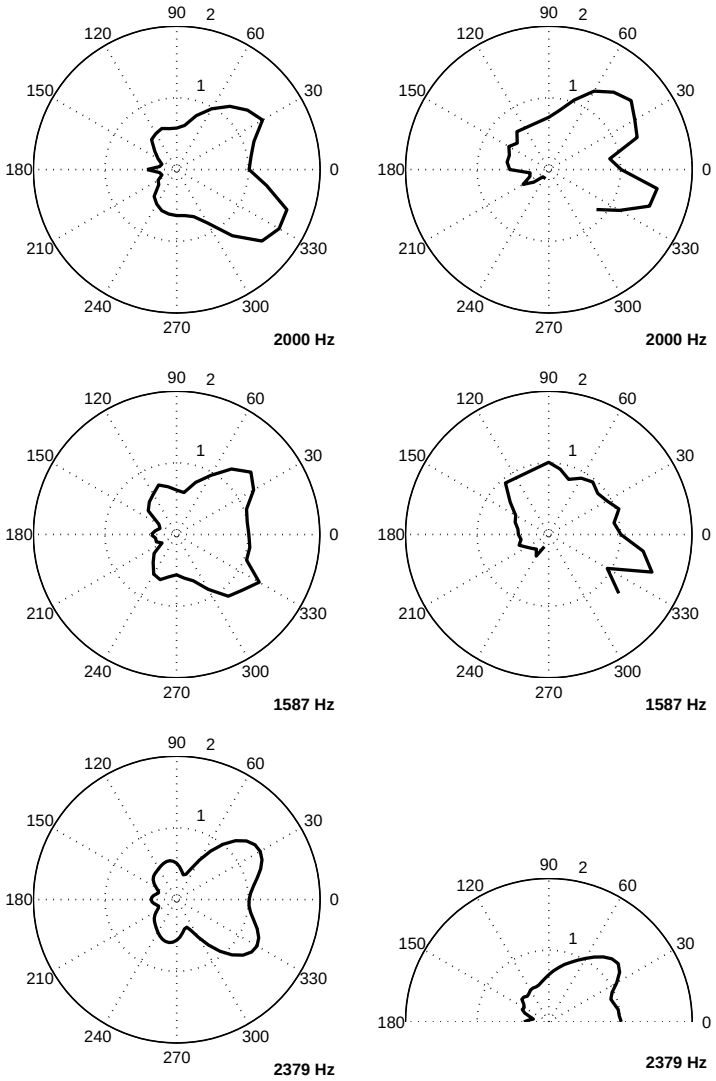


Figure 5.9: Comparison of directivity patterns of a female singer (top), a male singer (middle) and the artificial singer (bottom) around 2000 Hz

served in the sagittal plane: the male singer exhibits a less prominent sidelobe in the upper hemisphere compared to the female and the artificial singer.

5.3 Discussion

The radiation of a singer can be measured quite accurately with the described method. However, the time needed for such a measurement is rather long, and the procedure is exhausting for the singer under test. The radiation data from human singers and the data from the artificial singer are in good agreement. For experimental purposes the artificial singer can be employed as a realistic sound source when fed with a signal measured at the mouth of a human singer or originating from a voice simulation algorithm as described in this thesis.

Chapter 6

Singing voice synthesis

The origin of physical voice synthesis goes back to the speaking machine of W. Ritter von Kempelen in 1769 [vK91]. The machine was capable of producing vowels as well as consonants and even nasal sounds. An overview about the fascinating history of voice synthesis can be found in [Sch99]. More recently, the synthesis of singing voice has successfully been performed by sophisticated algorithms based upon a source-filter approach [Coo90], sampling techniques as available in musical synthesisers or a combination of both. A highlight in singing voice synthesis might have been the computer-assisted fusion of a male and a female voice for the sound track of the 1995 film *Farinelli – Il Castrato* at IRCAM, Paris.

However, these models – with exception of von Kempelen's physical model – are limited with respect to the possible interaction between the modules used for sound generation and cannot be called physical models. An approach that takes into account these interactions is not linear anymore but needs to allow signal flow between the modules.

6.1 Implemented model

For singing voice synthesis a combined model was used, consisting of separate models for vocal folds, vocal tract and noise generation as described in sections 2.3, 3.3 and 4.3.

The interaction of the models can be controlled in the graphical user interface with buttons that change the following parameters:

- Reflected flow from glottis into VT on/off
- Noise generation on/off

Future work will provide switches for the reflected flow from the VT into the glottis and the reflected flow from the lungs into the glottis.

The signal flow of the combined model is given in Figure 6.1. After initialisation of the global parameters the vocal fold signal is calculated sample-wise. The sound pressure wave is fed into the vocal tract model and the resulting pressure wave at the mouth is stored.

All user-defined parameters are modifiable during calculation and allow an arbitrary change of the boundary conditions.

Since the stability of the multiconvolution model has not allowed simulations of signals longer than about 30 ms (cf. section 3.4), the CTIM method has not yet been applied to the generation of continuous waveforms. Nevertheless, it has been used for the calculation of VTTFs as the length of the signals has not exceeded 15..20 ms.

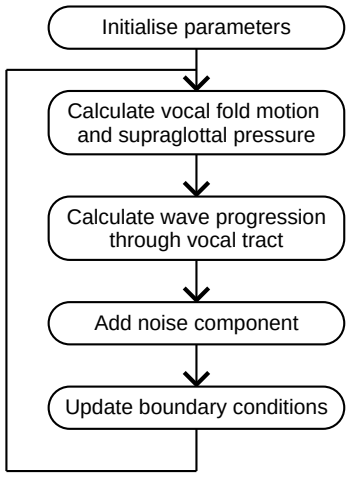


Figure 6.1: Signal flow of the combined modules

6.2 Interaction between vocal folds and vocal tract

The waveguide models described in chapter 3.3 assume propagation of a wave that originates from the flow modulation at the glottis. Due to the concept of superposition of two waves travelling in opposite directions, several ways of interaction between vocal folds and vocal tract can be considered.

The most important relation between both voice components is the pressure wave that enters the vocal tract. Without this component no phonation would be possible. A simple source-filter approach as described in the following section considers only this one-way relation.

The waveguide approach allows to take into account more complex interaction than the oscillator-filter model since the flow dependence on the input impedances of vocal tract and subglottal areas can be modelled. In 1837, the influence of acoustic impedances on the VF vibration has been studied by J. Müller [Mül37]. S. Hertegård and J. Gauffin examined the interaction between the voice source and the VT [HG93]. I.R. Titze reports that the subglottal resonance is important for register changes [Tit88]. Only in the two cases in which the vocal tract can be seen either as an infinite tube with the radius of the supraglottal area or as an acoustical swamp (no reflections), the input impedance has no influence on the flow transfer from glottis into the vocal tract.

6.2.1 Convolution

The most simple model for the synthesis of voice is the 'classical' oscillator-filter model that consists of a glottis model and a vocal tract model, which are independent of each other. For voice synthesis a vocal fold signal of arbitrary pitch is generated and convolved with a formant filter that represents the VTTF for a vowel sound.

An alternative approach has been described by M. M. Sondhi and J. Schroeter [SS87]. It is based upon a combination of a time-domain model for the vocal fold signal and a frequency-domain model for the vocal tract.

6.2.2 Vocal fold impedance

The basic assumption of the source-filter model is the unidirectional propagation of an acoustic wave from the glottis to the mouth. In the human voice organs some evidence exists for the need of a more complex model. The sound wave that is generated by the vocal fold movement propagates in two directions, upwards towards the mouth and downwards towards the lungs. As a first approach, the sound wave that travels into the lungs is absorbed with exception of low frequencies (DC component). The wave that travels upwards is not absorbed but rather reflected at constrictions within the VT and finally reflected to a significant part at the mouth opening. The reflected waves travel back towards the glottis and will there be reflected and transmitted again. As a result, a standing wave pattern is built inside the vocal tract that acts as a variant resistor for the incident sound wave from the glottis.

Glottis impedance

The reflection of acoustic waves at the glottis depends on the actual cross section of the glottis. Values from 1..10 mm^2 can be assumed for the glottal slit, corresponding to 1..30 % of the sub- and supraglottal areas. As a rough approach, the glottis can be said to be acoustically hard-walled for waves that are reflected there. Consequently, a big part of the energy that is directed towards the glottis will be reflected, and the reflected wave will interfere with the incident wave. The influence of different reflection coefficients at the glottis that take into account the change in glottal area during the oscillation cycle has been investigated by M. Liu and A. Lacroix [LL98]. Y. Pham Thi Ngoc states that the dependence of the glottis impedance on the glottal area can be described as a function of laminar losses, turbulent losses, and the impedance of a tube section of glottal area [Pha95].

In addition, the opening varies with the phase of the VF cycle, resulting in modulation of the reflection factors and therefore also in modulation of the standing

wave patterns and the formant structure of the VT. This phenomenon has been observed when comparing the formant structure of voiced and whispered speech [Jov98]. Impedance measurements of the vocal tract indicate significant changes of the VTMI for open and closed glottis (see section 3.5.2).

In Figure 6.2 the differences in the supraglottal pressure for different degrees of interaction between the vocal fold and the vocal tract model is depicted.

The dotted curve represents the supraglottal pressure wave generated without any interaction, the thin solid line indicates the course with time-variant glottal impedance enabled and the bold solid line represents the course with the glottal impedance and noise module switched on.

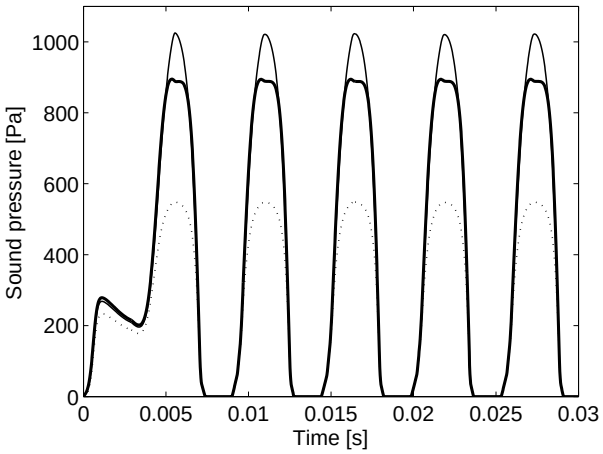


Figure 6.2: Comparison of different degrees of interaction

The supraglottal waveform is affected with respect to amplitude and ripple during the open phase. The period length is not changed. Perceptually, the difference between the waveforms was little.

6.3 Singing voice synthesis

The synthesis of the singing voice was carried out for configurations of the voice organ for vowels, overtone singing and two pathological cases.

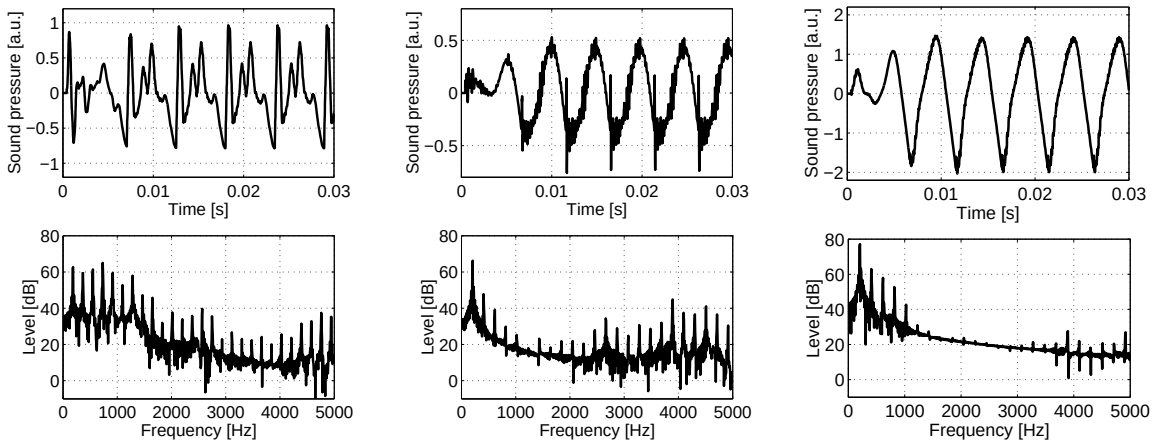
6.3.1 Vowels in modal, head and falsetto register

A satisfactory synthesis of vowels is the primary aim of this thesis. For inclusion of all relevant parameters and models, the combined model consists of the combination of the multiple-mass fold model attached to the waveguide vocal tract model. The radiation was calculated using the simple, low-frequency approach.

Modal register

For the calculation of the examples for voice in modal register the standard configuration as given in section 2.3.4 has been used. As examples, the German long vowels [a:], [i:] and [u:] are shown in Figure 6.3.

A comparison of the spectra of the different vowels illustrates the variety of the generated sound pressure signals.



(a) vowel [a:] (b) vowel [i:] (c) vowel [u:]

Figure 6.3: Time signal and spectrum of different vowels, modal register

Head register

The fundamental frequency of the head register voice is about 500 Hz for the following configuration:

- doubling of the *mucosa* tensions,
- division of the oscillating masses by a factor 10 due to lengthening and thinning of the VF tissues,
- reduction of the *vocalis* tensions $T_{v,act}$ and the the tension forces to a tenth of the standard value,
- increase of VF distance from the symmetry line to 0.175 mm.

From the simulated sound pressure it is obvious that the VF do not close completely. The remaining glottal gap yields a DC offset of the pressure signal. The spectrum should indicate a rather poor overtone structure. However, for the chosen configuration, the higher harmonics are still rather strong.

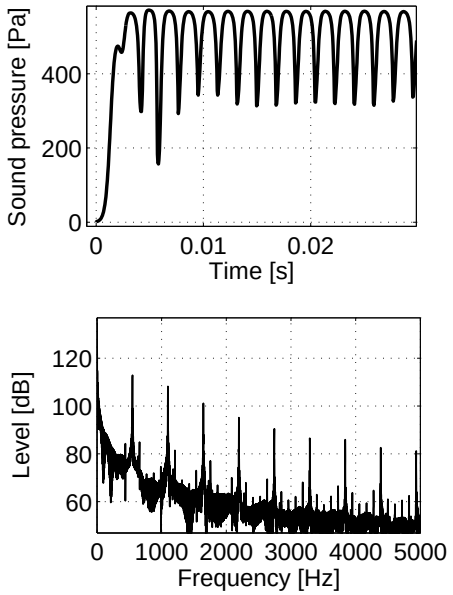


Figure 6.4: Time signal and spectrum of the supraglottal sound pressure, head register

Falsetto register

Male singers can extend the head register to higher frequencies with the falsetto register. Physiologically this mode of phonation is characterised by a reduction of the oscillating VF length due to a contraction of the *musculus lateralis*.

The falsetto register is not a completely new register but develops from the head register by modification of the following properties:

- reduction of the oscillating length of the VF to 2/3; within the model, this modification is achieved by a reduction of l_g to 2/3 of the standard value,
- reduction of the VF distance to a small value of 1 μm,
- reduction of the subglottal value to 2/3 of the standard value.

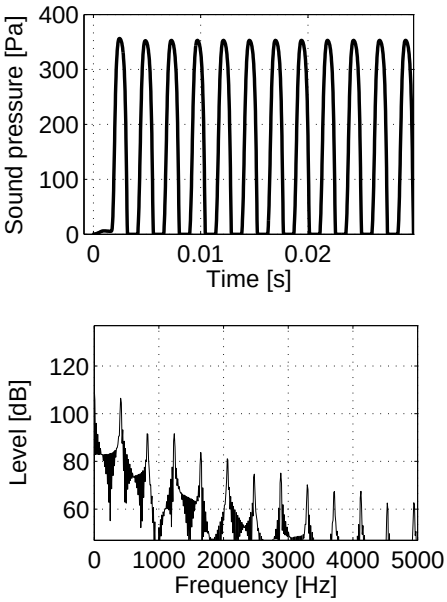


Figure 6.5: Time signal and spectrum of the supraglottal sound pressure, falsetto register

In Figure 6.5 the results from a simulation using the above configuration is depicted. The fundamental frequency of the generated sound is 350 Hz which is in the upper range of a Tenor singer. In contrast to the simulation of the head register, the glottis closes completely in the falsetto simulation.

Jitter and shimmer vs. vibrato

For a generation of jitter and shimmer or vibrato that is based on physical principles, different modes for modulation can be applied. As described in section 1.1.2, the origin of jitter, shimmer and vibrato can be manifold. The following features are considered in the implemented model.

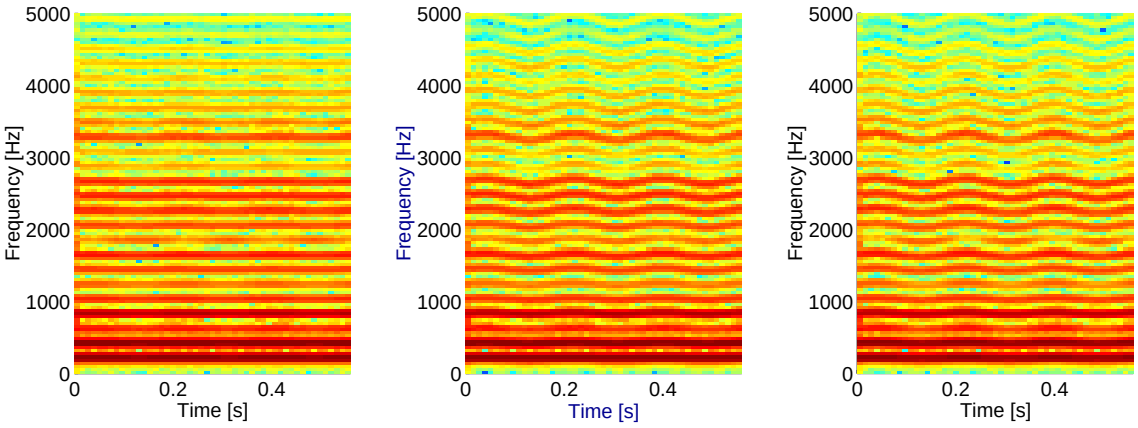
Glottal noise: The generation of aspiration noise will introduce small fluctuations of the sound pressure just above the glottis, as described in section 4.1.2. Due to its generation process, these fluctuations are chronologically correlated with the open phase of the glottal cycle (pulsed noise), but otherwise not controlled intentionally. Consequently, the modeling of glottal noise is not parametrised but switched on/off only in the GUI.

Subglottal pressure: One way to create vibrato would be the modulation of the subglottal pressure which corresponds to the breathing vibrato described in [Fis93]. The periodic movement of the *diaphragma* and the *peritoneum* can be summed up by the periodic modulation of p_{sub} with a frequency of 3.4..4 Hz. The vibrato can be controlled by periodic variation of the subglottal pressure with an arbitrary frequency and an amplitude given in percent of the DC lung pressure.

Vocalis stress: Adjustment of the (active) *vocalis* stress is the most important factor for the value of the fundamental frequency. A fast periodic modulation of the *vocalis* stress leads to jitter, a slow variation causes vibrato.

The two vibrato forms due to modulation of the subglottal pressure and *vocalis* stress have been simulated under the assumption of a synchronous, periodic modulation with 5 % modulation depth.

In Figure 6.6 spectrograms for three different kinds of vibrato modulation are presented. The breathing wave vibrato is shown in (a), the vibrato caused by modulation



(a) p_{lung} (b) $T_{v,act}$ (c) p_{lung} and $T_{v,act}$

Figure 6.6: Spectrograms for different vibrato modulations

of the *vocalis* tension is depicted in (b) and the spectrum resulting from a combination of both modulation methods is shown in (c). The effect of the breathing wave vibrato is rather weak compared to the results in (b) and (c) but clearly identified as vibrato by a listener. The result from the *vocalis* modulation was a rather profound vibrato that was amplified for case (c).

Supraglottal VT area: The entry area of the vocal tract is responsible for the first reflection of the glottis wave when leaving the glottis. The lower VT area (3..4 cm) just above the glottis has been found responsible for the singer's formant [Sun87].

Furthermore, video recordings of the supraglottal area of singers during phonation revealed a movement of the vocal tract walls a few centimetres above the glottis [NR01]. The results from simulation of moderate (5 %) change of the supraglottal EAF exhibit a weak, vibrato-like modulation of the supraglottal pressure wave.

6.3.2 Vocal fry

The synthesis of the vocal fry or 'Strohbaß' register has been performed with a fold configuration as for modal register except for the active stress $T_{v,act}$ that has been set to zero. Consequently, the vocal folds are driven by the forces in $x-$direction only, which corresponds to a very loose voice control.

Figure 6.7 illustrates the irregular cycle of the VF movement (pressure signal, top). The spectrum (bottom) exhibits subharmonics as observed in human vocal fry recordings. Note that the frequency axis is limited to 2 kHz for a better visualization of the low frequency components.

An increase of the active *vocalis* stress yields a more regular vibration pattern.

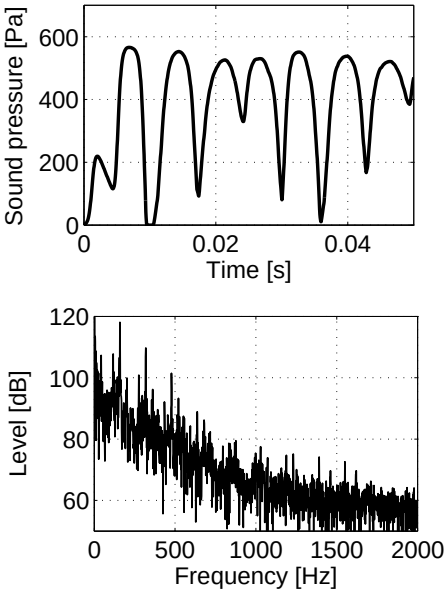


Figure 6.7: Time signal and spectrum of the supraglottal sound pressure, vocal fry

6.3.3 Overtone singing

Prior to the simulation results, some measurements and calculations are presented which made it possible to properly adjust the parameters of the vocal tract.

Measurements

Measured data of overtone singers is relatively rare. This might be caused by the fact that insight into the function of biphonic singing is of minor interest to most artists because the determination of voice physiology is mostly invasive or very costly (laryngoscopy, MRI). However, noninvasive sonographic and acoustic measurements are possible.

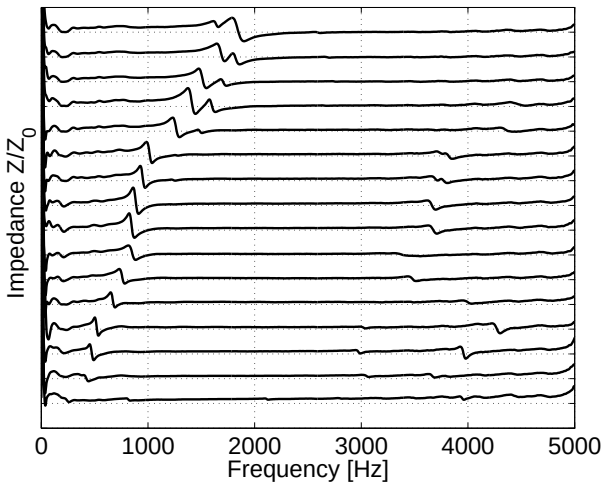


Figure 6.8: VTMI measurements of a biphonic sequence, rising melody pitch

In Figure 6.8 a sequence of VTMI measurements is shown that were obtained using the method described in section 3.5.2. The curves are shifted (from bottom to top) to visualise the course of time during the phonation of the rising overtone. The impedances are absolute values divided by the field impedance, i. e. the impedance with no vocal tract attached to the horn.

The plot illustrates that, apart from the overtone, only relatively weak resonances are excited between 3 kHz and 4 kHz. At higher resonances in the upper part of the plot a double resonance can be observed. This indicates that the overtone singer does not form a single resonance at the frequency of the melody tone but rather two closely neighboured resonances.

Figure 6.9 shows the results of another VTMI measurement on the same subject. The sequence of shifted curves demontrates the "morphing" from vowel [a:] (bottom) to the configuration of an overtone (top).

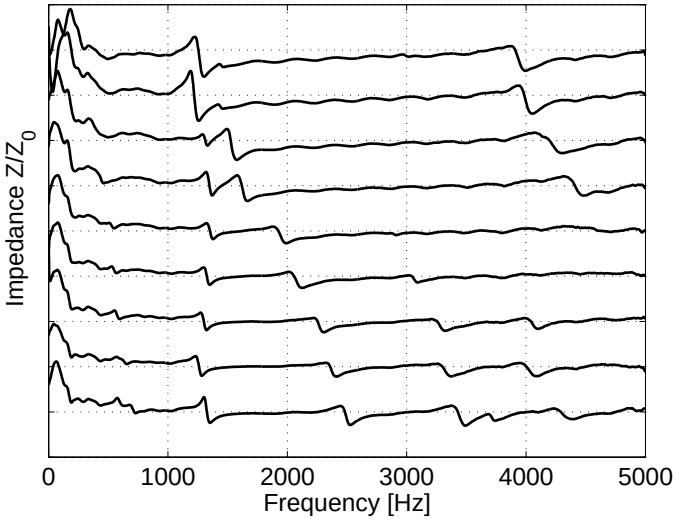


Figure 6.9: Sequence of VTMI measurements, morphing from vowel [a:] to an overtone

It is interesting to note that the second formant around 1300 Hz does not move significantly during the course of the sequence whereas the 3^{rd} formant moves from 2500 Hz downwards until it merges with the second one. All other frequencies are increasingly damped towards the overtone configuration. However, a weak resonance can be observed at 4 kHz.

The effect of formant merging is known as "focalisation" and can also be found in the transition from [y:] to [i:] as well.[1]

[1] P. Badin, personal communication

Calculations

The synthesis of overtone singing aims at a verification of the modelled process of sound generation. In a first step the vocal tract geometry is modelled using MRI data obtained by Adachi and Yamada [AY99] from a Xöömij overtone singer for four different melody pitches F_6, G_6, A_6, C_7 which equal 1397 Hz, 1568 Hz, 1760 Hz and 2093 Hz respectively. Figure 6.10 shows the equivalent area functions for these four configurations. In (6.1) the equations for calculation of the resonance frequencies

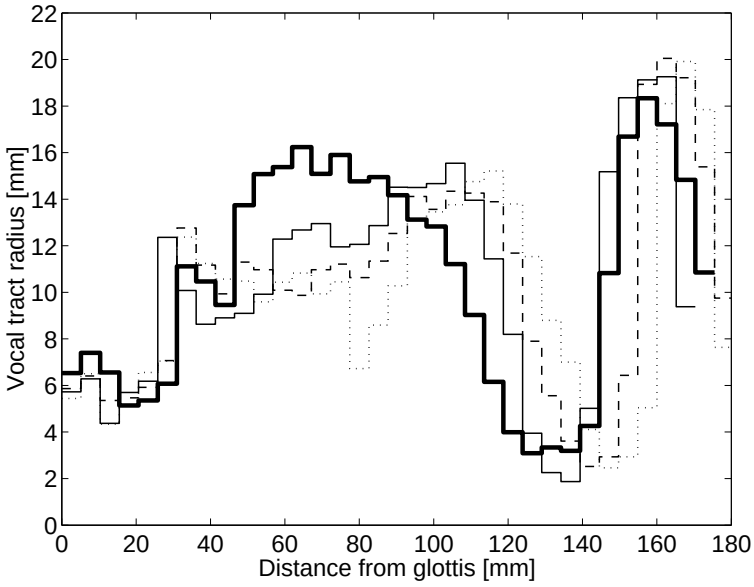


Figure 6.10: Vocal tract radius functions for overtones F_6 (dotted), G_6 (dashed), A_6 (thin solid), C_7 (thick solid)

are given. The first equation relates the resonance frequency to the length l_l of a $\lambda/2$ resonator and the speed of sound c_0. The assumption of an acoustically hard surface at the glottis and at the constriction generally results in too high values for the resonance frequencies but should be a first order approximation.

$$f_l = \frac{c_0}{2l_l}, \; f_H = \frac{c_0}{2\pi}\sqrt{\frac{S}{Vh}} \tag{6.1}$$

The second equation describes a Helmholtz resonator with the area of the mouth opening $S = \pi r_m^2$, the volume $V = \Delta x \pi \sum_i r_i^2$ of the Helmholtz resonator and the length of the mouth opening $h = \Delta x + 0.8\ r_m$. The last term $0.8\ r_m$ is a length correction for a circular opening.

In Figure 6.11 the comparison of the original and the modified area functions of the overtone A_6 as well as the corresponding vocal tract transfer functions (VTTF) are shown. The area functions taken from S. Adachi (thin solid line, visible at the mouth) have been modified at the mouth opening (thick solid line). The VTTF calculation has been carried out using the CTIM algorithm. An improvement of the second resonance by about 15 dB could be achieved by matching the resonance frequencies of the two resonators. Similar results can be obtained using the KL model, although the amplification is less impressive compared to the CTIM model.

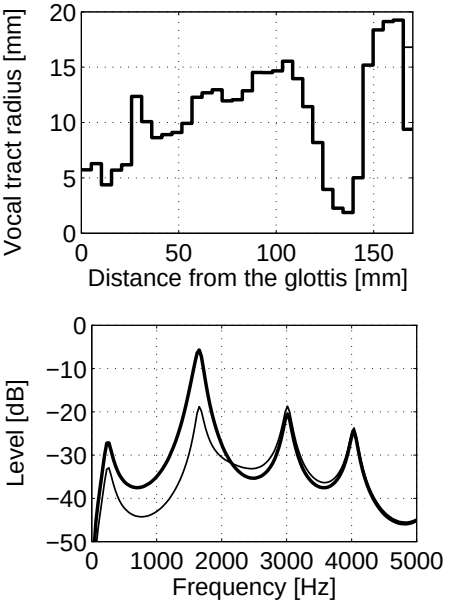


Figure 6.11: Area functions (top) and VTTF (bottom) of overtone A_6

Synthesis

In Table 6.1 a comparison of overtone resonance frequencies and levels calculated analytically from the above theory and from simulations using either the disc model or the cone model is presented. In the left column the resonance

Table 6.1: Comparison of calculated overtone resonances

		analytical		disk model		cone model	
Overtones after [AY99]		f_l	f_H	f_2	$P_{f_2} - P_{f_3}$	f_2	$P_{f_2} - P_{f_3}$
Pitch	[Hz]	[Hz]	[Hz]	[Hz]	[dB]	[Hz]	[dB]
F_6	1397	1304	3144	1325	-6.5	1378	3
G_6	1568	1356	2416	1550	2	1500	6
A_6	1760	1413	2616	1655	0.8	1570	7
C_7	2093	1541	2619	1960	2.5	1810	39
Optimized overtones							
F_6	1397	1304	1344	1325	15.5	1378	23
G_6	1568	1356	1760	1550	13	1500	25
A_6	1771	1413	1771	1655	16	1570	24.5
C_7	2093	1541	1986	1950	7	1810	10

frequencies using the original configuration in [AY99] are given. The second and third columns indicate the longitudinal and Helmholtz resonances calculated from equations (6.1).

The fourth and fifth column show the resonance frequencies and the sound pressure level difference between the second and third "formant" using the waveguide

model. The pressure level difference demonstrates the amount of damping that reduces the higher harmonics. The last two columns show corresponding results using the multiconvolution model.

6.3.4 Pathologic voice

The presented model can be applied to simulations of pathologic voice generation. Two different cases have been investigated in more detail: edema and vocal fold nodules. These diseases often occur in professional voice users such as singers or speakers. The modeling of other diseases such as lesions has not yet been implemented.

Edema

The diagnosis of an edema is present when at the inner side of the vocal folds an increased geometry and mass of the mucosa tissue is observed caused by an additional accumulation of liquid. Reasons for this pathology are often an inflammation and/or misuse of the voice.

The edema has been modelled by distributing a mass increase over about one third of the vocal fold, ranging from 1.1 to 1.5 times the specific *vocalis* mass and ranging from 1.5 to 3 times the specific mucosa mass. The set-up is shown in Figure 6.12. In

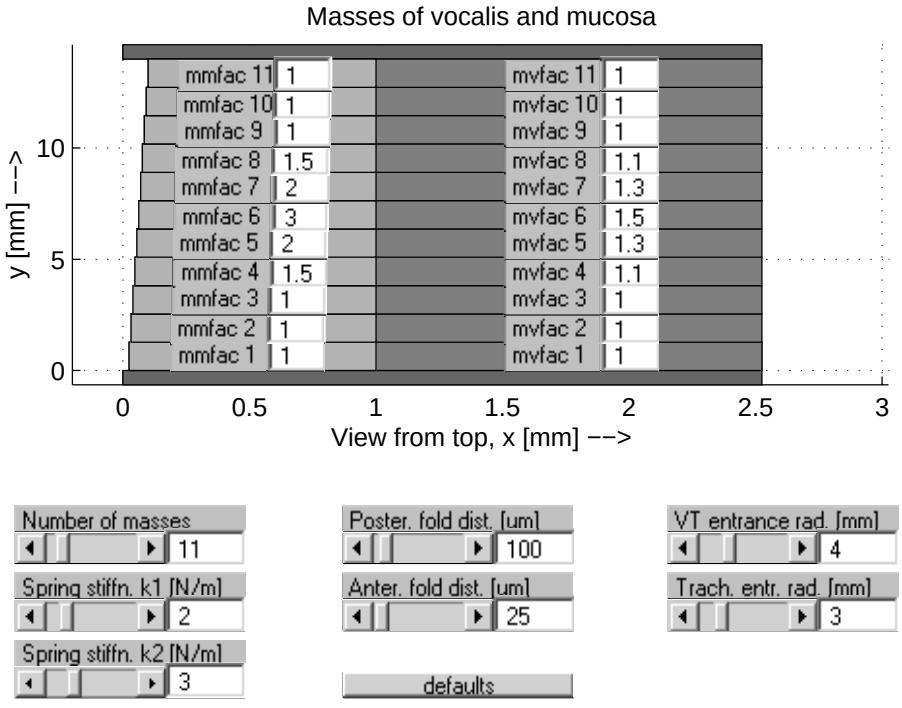


Figure 6.12: Mass distribution for fold movement with edema

Figure 6.13 the results of a simulation are given.

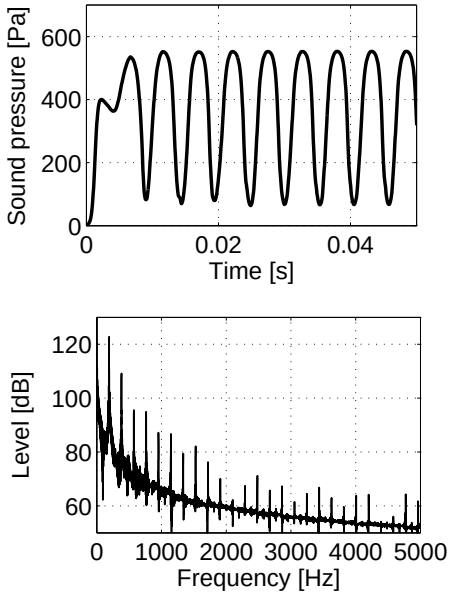


Figure 6.13: Supraglottal pressure for fold movement with edema

The results do not exhibit any particular change compared to a healthy voice (cf. Fig. 2.17 on page 38) with exception of a lowered fundamental frequency due to the increased mass.

Fold nodules

Singer's nodules are characterised by a local increase of tissue, in most cases on opposite locations of the vocal folds. In Figure 6.14 an example of such a singer's nodule is pictured [Kle99].

Figure 6.14: View upon vocal folds with singer's nodules

The set-up for the calculation of the singer's nodule is similar to the set-up for the edema, with the exception that only one mass of one segment out of 11 vocal fold segments is increased. A factor of 3 has been chosen, both for vocal fold masses and the mucosa masses.

The observation of the animation of the masses reveals an irregular VF movement that begins regularly with the first (1,0) mode. After a few cycles, the oscillation shifts to the second (longitudinal) (2,0) mode and stays in a periodic but rather complex movement pattern. The fundamental frequency is divided by two

compared to the simulation of a healthy voice. The sound pressure and the spectrum of the resulting signal are depicted in Figure 6.15.

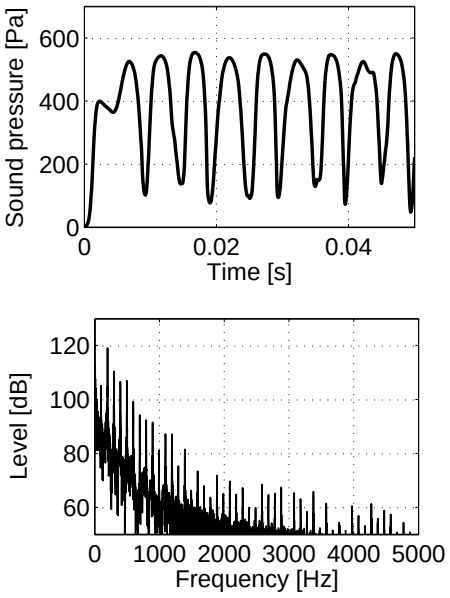


Figure 6.15: Supraglottal pressure for fold movement with nodule

The signal exhibits a significantly lower fundamental frequency that is due to the irregular VF movement pattern in the (2,0) mode.

In contrast to the simulations the *in vivo* observation of the vocal folds in patients suffering from singer's nodules does not necessarily exhibit an irregular VF movement. This might be caused by an automatic compensatory parameter change. It was found that the irregular movement can be turned into a regular one by increasing the active stress $T_{v,act}$ on the *vocalis* muscle. With increasing $T_{v,act}$, the fundamental frequency rises and, at a certain value, the oscillation falls back to the first (1,0) mode. It was discovered that by successively increasing $T_{v,act}$, the change between the two states of oscillation takes place when the active stress is increased by ca. 5 %. A spectrogram of the transition is shown in Figure 6.16.

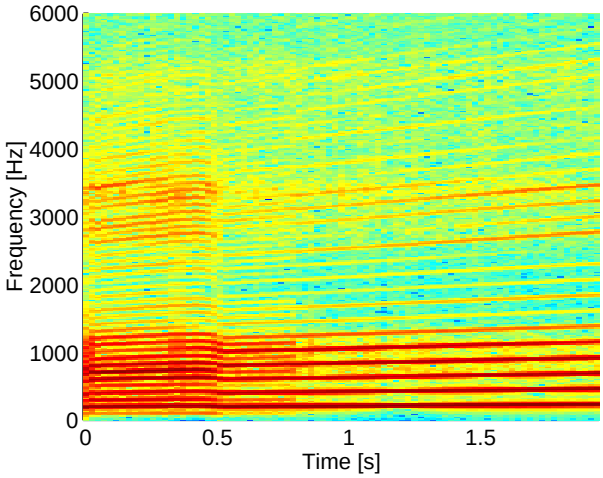


Figure 6.16: Spectrogram for the transition from second to first mode

In the Figure the transition of the (2,0) mode into the (1,0) mode can be observed at the time 0.5 seconds.[2]

[2]Sound examples for most simulations can be found on the internet page `http://www.akustik.rwth-aachen.de/~malte/vox`

6.4 Discussion

As presented in this chapter, the task of simulating the singing voice has been successfully carried out for selected configurations. The parameters were chosen as similar as possible to actual parameters within the human voice organ.

Future extensions of the vocal fry simulations with an asymmetric VF model could lead to a deeper understanding of the aerodynamic coupling of both vocal folds. The extension of the present multimass model into an asymmetric model could be useful for the modeling of biphonic voice sounds since asymmetric models have already been successfully applied to the simulation of pathologic voice [AB01, SH95].

The vocal tract models seem to fulfill the requirements even for extreme VT configurations as indicated by the simulation of overtone singing.

In the case of the simulation of pathologic voice, it is not easy to derive quantitative rules for the description of pathologies from the results, since both main modules, vocal fold model and vocal tract model, still need to be improved for more realistic voice synthesis. Nevertheless, it could be demonstrated that the adaption of vocal fold parameters for the case of a singer's nodule yields reasonable results. Moreover, the effect of a compensation of the nodule by increase of the *vocalis* tension has also been successfully modelled.

The investigation of the interaction between vocal tract and vocal folds is still progressing. The wave that is reflected back into the glottis should also be taken into account for the calculation of the pressures within the vocal fold model. Preliminary results, however, indicate that the pressure wave from the vocal tract drastically changes the supraglottal pressure conditions in a way that oscillation cannot be sustained. Further research must be carried out to find the reason for this behaviour.

Chapter 7

Summary, conclusions and outlook

7.1 Summary

To understand the physics of the human voice is a challenge that requires knowledge of several different scientific fields. For a correct description of the boundary conditions, insight into geometry and biomechanical construction of lungs, larynx and vocal tract is necessary.

The voice generation process can be divided into several discrete models. In this work appropriate mathematical models for simulation of vocal fold movement, wave propagation through the vocal tract and noise generation at the glottis have been implemented. All code has been written in MATLAB to provide easy visualization of dynamic changes in geometry and auralisation of the generated signals.

The vocal fold model is a symmetric multiple mass model that consists of an arbitrary number of two-mass-segments for each fold. The aerodynamics are modelled by assumption of a free jet that separates from the vocal folds.

A noise module models the creation and propagation of *vortices* and noise generation due to turbulences in dependence of flow parameters and dynamic geometry changes of glottis and vocal tract.

Sound propagation through the vocal tract is modelled either by the classical waveguide algorithm or by a multiconvolution technique that allows arbitrary sample rates and less computational effort. The losses due to radiation have been modelled either with a low-frequency approach or by use of digital filters that simulate a baffled piston.

Furthermore, the radiation of an artificial singer has been investigated. A replica of an adult human's head and upper torso has been built. The "hummy head" has been realised by insertion of loudspeaker systems into the replica of a human torso. A comparison of directivity measurements of human singers showed good agreement with the radiation characteristic of the artificial singer.

The simulations with the combined model, consisting of a vocal fold model, a

vocal tract model and the radiation, reveal a mutual dependence of the components. Therefore, synthesis of a naturally sounding voice cannot be done with the simple linear oscillator-filter model but rather with a combined model that takes into account pressure flows and connection impedances between models. The naturalness of synthesised vowels could be significantly improved.

The combined model has been applied to the synthesis of overtone singing. Sonographic imaging and analytical calculations based on magnet resonance imaging (MRI) data proved that the melody pitch during overtone singing is caused by superposition of two resonances. The first resonance is related to a longitudinal $\lambda/2$ resonance between glottis and a constriction formed by the tongue whereas a Helmholtz resonator between the constriction and the mouth opening causes the second resonance.

Results of the voice simulation have been compared to measurements of the sound pressure at the glottis. Additional *in situ* measurements were carried out for the verification of the pressure transfer function calculations of the vocal tract. A novel method for measurement of the vocal tract mouth impedance (VTMI) has been applied to the determination of the vocal tract resonances. The application of this method allows a non-invasive analysis of different configurations of the voice organ and provides a sufficiently good signal-to-noise ratio for measurement of the vocal tract resonances during phonation. An application of this method for the diagnosis and therapy of articulatory dysfunction is described. The feasibility of the method in clinical use is subject to current research.

7.2 Conclusions

This work presents a physical model for voice synthesis that serves both as a research device and as an educational tool for modeling, visualization and auralisation of the human voice. The challenge of building such a model is to find a trade-off between a detailed description of the physiology and the physical relations that are involved on one hand and, on the other hand, the implementation of fast and stable algorithms. Due to the complexity of the voice generation process and the limited time within the frame of a doctoral thesis, only the most important models for generation of sustained vowels have so far been implemented. The focus was laid on the mathematical description of the main components of voice generation with an emphasis on singing voice synthesis. As a consequence, no attempt has been made to “improve” the generated signals by means of post-processing techniques such as filtering to obtain a more realistic sound.

The main topics of the thesis are modeling of

- sound generation due to the vocal fold movement,
- noise generation at the glottis,
- sound propagation through the vocal tract, and
- radiation characteristics of the a singer.

For all models, algorithms described in literature were modified, implemented in MATLAB and checked with respect to their suitability for singing voice synthesis. Modifications have been applied to the models to include latest developments.

These models were coupled in several ways. The influence of their coupling has been investigated and it could be shown that the simple oscillator-filter approach is of limited accuracy. The combined model includes noise generation that is based upon the vortex sound generation theory. The noise is inserted correctly into the vocal tract. Furthermore, new measurement techniques have been developed to verify the simulation results. The methods include

- direct measurement of the vocal tract transfer function (VTTF),
- measurement of the vocal tract mouth impedance (VTMI),
- determination of the harmonics-to-noise ratio (HNR), and
- measurement of the directivity of a singer.

The VTTF measurements yield correct results for the determination of the vocal tract resonances, but the measurement procedure is rather invasive and not suitable for investigations in most singers. However, results from measurements of the VTTF can be used for characterization of individual articulatory configurations. The method has been successfully applied on one subject and a comparison of the results to simulations and literature indicated a good agreement of the measured frequencies within the expected range.

The VTMI method is a non-invasive way to characterise the vocal tract resonances. A high correlation between resonance frequencies and literature or VTTF measurements has been found. Due to the external excitation of the vocal tract, phonation of the singer under test is not necessary. As a consequence, the medical application of the VTMI method to patients with no or weak phonation is superior to classical measurement methods such as LPC analysis.

Some problems still remain to be investigated more in detail. Since the main idea of a physical model is the description based on exact input data or at least physically realistic data, the realism of results from such models depends on their complexity and the accuracy of the parameters used.

7.3 Future development

The future development of physical voice models and measurement methods could evolve in several directions.

Aerodynamics

One of the most important topics with respect to a correct description of the voice generation process would be the investigation of the aerodynamic flow within and above the glottis during phonation. The time- and space-variant determination of the jet separation point seems to be crucial for the calculation of the aerodynamic forces on the vocal folds [PHWB95]. A continuous 3-dimensional description for the vocal fold surface could combine the multimass model described in this thesis with the aerodynamic models of Lous [LHVH98] and Vilain [VPH+01].

The application of miniature flow sensors such as the microflown could yield information about the flow distribution during *in vivo* measurements. However, the protection of the sensor against the humid environment is problematic.

Asymmetric fold model

An extension of the model towards asymmetric vocal fold configurations is in fact possible. Investigations of the effect of asymmetry have been done by Steinecke and Herzel [SH95] and Tigges [TMH+97].

Alternative models

A different approach would be the modeling of the vocal fold tissue or the vocal tract geometry with finite elements. Recently, Alipour [AB01] demonstrated a 2-dimensional finite element model of the vocal folds. The calculation of waveforms is based upon a finite element model consisting of some hundreds of elements for each VF. However, the calculation of the aerodynamics for such a model is very complex because the geometry changes between each sample in direction of all co-ordinates.

Improvement of the implemented models

The following improvements should make the model more useful:

- Stabilization of the cone model
  The development of a numerically stable algorithm for the CTIM calculation of the impulse response should reduce the computational costs and segments needed for articulatory synthesis.

- Generation of ventricular fold phonation
  More sophisticated voice generation mechanisms require the inclusion of the ventricular folds into the present model.

- Simulation of the singer's formant
  Variations of the EAF should enable the generation of a singing voice that exhibits the characteristic spectral enhancement known as singer's formant.

Medical applications

The preliminary results presented in section 6.3.4 indicate that distributed mass VF model is suitable to simulate pathologic voice production mechanisms such as those regarding edema or vocal fold nodules. It has also be shown that compensation mechanisms, for example increase of the *vocalis* tension can be modelled. Future applications of the models should extend the range of applications towards secondary effects of the primary pathology.

A more comprehensive model could include the morphology of exterior muscles that support the larynx such as *musculus sternothyreoideus*. The implementation of such an exterior muscular frame allows a simulation of mechanisms of physiologic and pathologic balance between interior and exterior muscles. Such a complex model could be used to study the interaction of interior muscular dysfunction and pathologies of external muscles strains. A clinical application could be the visualization and auralisation of such complex pathological findings.

An interesting extension of the VF model would be a more realistic representation of the surfaces. Texture mapping from photos of human VF tissue could render the model more realistic. Simulations of VF pathologies could then more easily be compared by doctors with video images of the vocal folds.

The implementation of morphologic and functional glottal asymmetries will allow a simulation of unilateral VF paralyses. Also, biphonic voice generation caused by different vibrational modes of each VF could be modelled using this asymmetry feature.

Finally, the demonstration of physiologic and pathologic mechanisms in teaching and consultation of patients will be an important application of the model. Once the model parameters are adapted to the patient's anatomy and physiology, a simulation of the individual voice characteristics could be achieved.

Chapter 8

Kurzfassung

Die Nummerierung der Abschnitte in dieser Kurzfassung entspricht der Nummerierung der Kapitel der Dissertation. Wegen der Kürze des Textes ist auf die Darstellung von Bildern verzichtet worden, daher sei zur Erläuterung der Abschnitte auf die Darstellungen in den entsprechenden Kapiteln der englischen Version verwiesen.

Einführung

Das Verständnis der physikalischen Ursachen der menschlichen Stimmerzeugung ist eine Herausforderung, die Kenntnisse in verschiedenen wissenschaftlichen Bereichen erfordert. Um die Vorgänge adäquat zu beschreiben und die Parameter realistisch zu konfigurieren ist das Verständnis der geometrischen und biomechanischen Zusammenhänge des Aufbaus der Lunge, des Kehlkopfs und des Ansatzrohrs nötig. Die Stimmerzeugung kann zur Beschreibung dieser Zusammenhänge in verschiedene Modelle zerlegt werden. In dieser Arbeit wurden jeweils mathematische Modelle für die Simulation der Stimmlippenbewegung, die Schallausbreitung durch das Ansatzrohr und die Rauscherzeugung im Kehlkopf geschaffen. Die Programmierung aller Module erfolgte in MATLAB, wodurch eine einfache Darstellung der Schwingungsbewegungen sowie eine Klangausgabe der simulierten Schallwellen möglich ist.

Zur Überprüfung der Simulationsergebnisse wurden z. T. eigene Methoden entwickelt, mit denen neue Erkenntnisse über die Funktion aussergewöhnlicher Gesangstechniken wie dem Obertonsingen gewonnen wurden.

8.1 Der Sänger

In diesem Kapitel werden die Signaleigenschaften der Singstimme beschrieben. Durch Steuerung der Stimmlippenfunktion ist der Sänger in der Lage, verschiedene Stimmregister zu wählen, durch die nicht nur der Stimmumfang, sondern auch die Klangfarbe

des Primärschalls bestimmt wird, d. h. die Zusammensetzung der erzeugten Obertöne. Die Geometrieänderung des Raumes zwischen Stimmlippen und Mundöffnung, des Ansatzrohrs, bestimmt jedoch in erster Linie den Klang des abgestrahlten Schalls. Mit den wichtigsten klangbestimmenden Artikulatoren Zunge, Kiefer und Lippen werden Laute gebildet, die im musikalischen Kontext mit oder ohne textliche Bindung als Melodie, Begleitstimme oder Chorklang zum Hörer gelangen.

Der komplexe Klang der menschlichen Stimme weist viele Eigenschaften von Musikinstrumenten auf, die einen angehaltenen Klang erzeugen. Zu diesen Charakteristika gehört eine regelmäßige Obertonstruktur, ein nichtlinearer Oszillator für die Primärschallanregung (Stimmlippensignal), die Klangformung durch einen angekoppelten Resonator, eine große Variabilität der Grundfrequenz sowie Mikroschwankungen der Signalamplitude (shimmer) und Momentanfrequenz (jitter). Die Besonderheit des Chorklangs wird anhand der Kohärenz stimmhafter und stimmloser Stimmklänge erläutert. In der Diskussion wird die physikalische Modellierung der Stimmorganfunktionen den in der Sprachsynthese üblichen Verfahren zur Nachbildung des Stimmsignals gegenübergestellt.

8.2 Stimmlippen

Die Stimmlippen sind ein inhomogenes, anisotropes Gewebe, dessen Eigenschaften sowohl willkürlich zur Einstellung verschiedener Register als auch unwillkürlich während der Phonation in weiten Bereichen veränderlich sind. Ein Modell, das diese Eigenschaften exakt nachbildet, ist heute noch nicht verfügbar. Zur Annäherung der wirklichen Geometrie und Schwingungsform werden daher Modelle eingesetzt, welche lediglich die wesentlichen Funktionen nachbilden.

Stimmlippenmodelle

Das am weitesten verbreitete Stimmlippenmodell wurde 1972 von K. Ishizaka und J. L. Flanagan als gekoppelte Anordnung zweier Massen für eine Stimmlippe konstruiert. Es wird im Folgenden als IF-Modell bezeichnet. Das Modell bildet nur eine Stimmlippe nach; die Bewegung der zweiten Stimmlippe wird als symmetrisch zur Mitte der Glottisöffnung angenommen. Die beiden Massen werden mit unterschiedlichen Werten für Masse und Federkonstanten zur Anbindung an den angrenzenden Knorpel versehen, was eine Zuordnung der Massen zu den Geweben *vocalis*-Muskel und *mucosa* (Schleimhaut) nahe legt. Der Druckverlauf vom subglottischen Raum (Lungendruck) bis zum supraglottischen Raum (Druck im unteren Bereich des Ansatzrohres) ist in mehrere Sektoren unterteilt, in denen jeweils unterschiedliche Mechanismen für den Druckverlauf zugrunde gelegt werden (Pressure recovery, Coanda-Effekt).

Neuere Modelle setzen für die Schwingungserzeugung der Stimmlippen die Erzeugung eines Jets voraus, durch den während der Öffnungs- und Schließphase des Schwingungszyklus unterschiedliche Kräfte auf die Massen wirken. Hierdurch ist die Annahme der nicht nachgewiesenen Mechanismen des IF-Modells nicht mehr nötig, da die durch die Jetbildung verursachten Bernoullikräfte eine entsprechende Funktion übernehmen. Weitere Ansätze repräsentieren die anisotrope Gewebestruktur der Stimmlippen durch mehr als zwei Massen (Body-Cover-Modell, B. Story und I. R. Titze 1995).

Implementierung

Für die Nachbildung der Stimmlippen wurde ein symmetrisches Mehrmassenmodell geschaffen, das aus einer beliebigen Anzahl von Zwei-Massensegmenten für jede Stimmlippe besteht. Grundlage für diese Implementierung ist das 1973 von I. R. Titze entwickelte 16-Massen-Modell. Die aerodynamischen Eigenschaften wurden – im Unterschied zu Titzes Implementierung – unter der Annahme der Ablösung eines Jets von den Stimmlippen berücksichtigt. Die Unterteilung der Stimmlippe erfolgt in eine beliebige Anzahl von Segmenten, wodurch eine Nachbildung vom 2-Massen-Modell bis zu einer quasi-kontinuierlichen Beschreibung der Stimmlippen in Längsrichtung möglich ist. Als Ablösepunkt für den Jet wurde der Übergang zwischen den beiden Massen gewählt. In Abhängigkeit der relativen Stellung der beiden Massen je Segment wirkt die zusammenziehende Bernoullikraft nur auf den unteren, stromaufwärts befindlichen Teil der *vocalis*-Masse.

Simulationen

Über eine grafische Benutzerschnittstelle (graphical user interface, GUI) kann die Grundeinstellung für ein Register gewählt werden sowie physikalische Größen wie Lungendruck, Stimmlippenspannung, Öffnungsgrad und -winkel etc. eingestellt werden. Die Bewegung der Stimmlippen, die Werte der auf die einzelnen Massen wirkenden Kräfte und die aus der Berechnung resultierenden Größen Volumenfluss und supraglottaler Schalldruck können während der Berechnung visualisiert werden. Seitenansicht, Draufsicht und eine dreidimensionale Darstellung einer Stimmlippe sind simultan möglich.

Stimmlippensignale wurden für verschiedene Registereinstellungen berechnet. Die Simulationsergebnisse zeigen für Brust-, Kopf-, Falsett- und Strohbassregister eine gute Übereinstimmung mit perzeptiven Eindrücken gesungener Klänge. Diese Resultate decken sich auch mit Literaturangaben zu vergleichbaren Simulationen.

8.3 Ansatzrohr

Bei der Formung unterschiedlicher Klänge wird die Form des Ansatzrohres in weiten Grenzen verändert. Sowohl die Länge als auch der Verlauf der Querschnittsflächen als Funktion der Distanz von der Glottis sind für jeden erzeugten Klang unterschiedlich und verändern sich während des Singens und der Artikulation sehr stark.

Für die Modellierung der Wellenausbreitung im Ansatzrohr wird meist die Näherung einer ebenen Welle angenommen, wenn auch die Voraussetzung dieser Annahme für Frequenzen oberhalb von ca. 4 kHz nicht mehr für alle Ansatzrohrkonfigurationen gegeben ist. Da jedoch die wesentlichen klangbildenden Ansatzrohrresonanzen (Formanten) unterhalb dieser Frequenz liegen, wurde auch in dieser Arbeit eine eindimensionale Wellenausbreitung vorausgesetzt.

Implementierung

Die Schallausbreitung durch das Ansatzrohr wird entweder mit einem einfachen Wellenleiteralgorithmus oder einer Mehrfach-Faltungstechnik (continuous-time interpolated multiconvolution, CTIM) modelliert. Das CTIM-Verfahren bietet eine größere Freiheit bei der Wahl der Abtastrate und ist weniger rechenintensiv als das Wellenleiterverfahren.

Beide Berechnungsverfahren verwenden für die Nachbildung der Ansatzrohrgeometrie so genannte äquivalente Flächenfunktionen (equivalent area functions, EAF), die für Klänge unterschiedlicher Sprecher bzw. Sänger aus Magnet-Resonanz-Aufnahmen gewonnen wurden und in der Literatur verfügbar sind.

Die Modellierung der Schallabstrahlungsverluste bei der Ansatzrohrberechnung wurde auf zwei Arten realisiert; eine Methode nutzt einen einfachen, linearen Ansatz für die Beschreibung der Abstrahlung, die zweite Methode nähert die Abstrahlung an die eines Kolbenstrahlers in einer Wand an.

Simulationen

Für die Berechnung der Wellenausbreitung im Ansatzrohr wurde als Anregung an der Glottis ein Dirac-Impuls verwendet. Iterativ werden beim Wellenleiteralgorithmus für eine gegebene EAF-Konfiguration die hin- und die zurücklaufenden Wellen entlang des Ansatzrohres berechnet. Beim CTIM-Verfahren werden zunächst die Reflexionsfunktionen berechnet; erst dann wird die iterative Berechnung der Wellenausbreitung durchgeführt.

Die Programmoberfläche des Ansatzrohrmoduls erlaubt eine Variierung der EAF während der Berechnung, um Übergänge zwischen Lauten nachzubilden oder den Einfluss verschiedener Veränderungen der Artikulation zu simulieren.

Beide Verfahren ergeben für die aus der Literatur entnommenen EAF-Daten eine gute Übereinstimmung der Formantfrequenzen bis ca. 5 kHz. Der Verlauf der Verstärkung bzw. Dämpfung mit der Frequenz weist jedoch stärkere Abweichungen auf, deren Ursache vermutlich in der unzureichenden Berücksichtigung der frequenzabhängigen Dämpfungsmechanismen zu finden ist.

Messungen

Um eine Vergleichsmöglichkeit zwischen Simulationsergebnissen und den Vokaltraktkonfigurationen menschlicher Sänger zu schaffen, wurden mehrere Messverfahren verwendet. Einerseits wurde die Übertragungsfunktion des Ansatzrohres (vocal tract transfer function, VTTF) bestimmt, indem mit zwei Mikrofonen während externer oder interner breitbandiger Anregung der Schalldruck simultan an der Glottis und am Mund aufgezeichnet wurde. Eine weitere Methode wurde entwickelt, um die Impedanz des Ansatzrohrs am Mund (vocal tract impedance at the mouth, VTMI) zu bestimmen. Hierzu kam ein neuartiger Sensor zum Einsatz ($\mu-$flown), mit dem die Schallschnelle gemessen werden kann. Mit Hilfe dieser Methode können die Resonanzfrequenzen des Ansatzrohres bestimmt werden, auch wenn keine Phonation stattfindet. Der Einsatz der VTMI-Methode bei der Bestimmung der Artikulation beim Obertonsingen gab Hinweise auf den Funktionsmechanismus dieser speziellen Gesangstechnik.

8.4 Rauscherzeugung

Neben dem aus Harmonischen bestehenden Primärschall, der durch glottale Modulation des Volumenstroms erzeugt wird, gibt es im Stimmorgan auch Rauschquellen, die ihre Ursache in turbulenter Strömung haben. Es lassen sich zwei Grundarten von Rauschen unterscheiden: zum einen das Friktionsrauschen, das – ähnlich dem Schneidenton einer Lippenpfeife – bei der Anströmung einer Kante entsteht, und das Aspirationsrauschen, welches bei der Glottisbewegung selbst entsteht. Beide Rauscharten verhalten sich sehr ähnlich, doch tritt das Friktionsrauschen meist als eigener Klang auf (z.B. bei den Lauten [ʃ] oder [s]), während das Aspirationsrauschen bei der gesunden Stimme mit dem harmonischen Signal perzeptiv verschmilzt.

Implementierung

Für die Nachbildung des Friktionsrauschens wurde ein Ansatz nach D. J. Sinder (1999) implementiert und modifiziert. Ein Rauschmodul bildet die Entstehung und Ausbreitung von Wirbeln sowie die Erzeugung von Rauschen durch Turbulenzen nach.

Als Eingangsparameter für die Modellierung dienen der Fluss durch die Glottis und die dynamischen Geometrieänderungen von Glottis und Ansatzrohr. Das generierte Rauschsignal wird am Ort der Entstehung in die entsprechenden Segmente des Ansatzrohrs zu den von der Stimmlippenbewegung direkt erzeugten Schallwellen addiert.

Messungen

Die Erzeugung von Aspirationsrauschen beim Menschen wurde anhand einer Gruppe von elf unerfahrenen Sängern und zwei Sängern mit Chorerfahrung verifiziert. Dabei konnte ein Zusammenhang zwischen der Geometrie des supraglottalen Raumes im Ansatzrohr und dem HNR (harmonics-to-noise ratio) bzw. dem GNE-Verhältnis (glottal-to-noise excitation) bei verschiedenen Vokalen nachgewiesen werden. Zur Bestimmung des HNR wurde ein Implementierung des PSHF-Algorithmus von P. J. B. Jackson (1998) verwendet, das GNE-Verhältnis wurde mit dem Göttinger Heiserkeitsdiagramm (Michaelis 1997) bestimmt. Bei den Vokalen [o:] und [u:] konnte ein relativ geringer Rauschanteil gemessen werden, was mit einem sanft ansteigenden Verlauf der EAF korreliert. Bei den Klängen [æ:] und [e:] tritt hingegen ein relativ starker Rauschanteil auf; zugleich ist ein deutlich abrupterer Verlauf der EAF im Ansatzrohr unmittelbar oberhalb der Glottis beobachtbar.

8.5 Abstrahlung

Der Übergang des Ansatzrohrs in das Freifeld an der Mundöffnung hat sowohl Auswirkung auf die Klangbildung im Ansatzrohr als auch auf die Charakteristik der Abstrahlung in die Umgebung des Sängers. Während die Diskontinuität der akustischen Impedanz an der Mundöffnung entscheidend den Anteil der rücklaufenden Welle im Ansatzrohr und des ins umgebende Schallfeld transmittierten Schalls bestimmt, ist die Richtwirkung des Sängers durch die spezielle Geometrie des Kopfes und Oberkörpers sowie der Mundform bestimmt.

Es wurde die Schallabstrahlung eines künstlichen Sängers untersucht, der eine möglichst realistische Nachbildung des Oberkörpers und Kopfes eines erwachsenen Menschen darstellt. Dem künstlichen Sänger wurde durch den Einbau eines 2-Wege-Lautsprechersystems eine Stimme verliehen. Ein Vergleich der Richtcharakteristiken von menschlichen Sängern mit der des künstlichen Sängers zeigt eine generelle Übereinstimmung, doch sind sowohl Unterschiede in der Frequenzabhängigkeit zwischen verschiedenen Sängern als auch zwischen menschlichem und künstlichem Sänger zu beobachten.

8.6 Synthese der Singstimme

Die Nachbildung des Singstimmsignals auf Grundlage eines physikalischen Modells ist bereits im 18. Jahrhundert in Form einer sprechenden Maschine vorgenommen worden (von Kempelen, 1769). In neuerer Zeit sind zahlreiche Versuche unternommen worden, die Singstimme künstlich zu erzeugen, doch lassen die Ergebnisse bislang Natürlichkeit oder Flexibilität der register- und sängerspezifischen Eigenschaften vermissen. Die hier vorgestellte Implementierung legt den Schwerpunkt auf eine möglichst variable Anpassung des Modells an eine gegebene Physiologie.

Implementiertes Modell

Das Modell zur Singstimmsimulation besteht aus den in den vorangegangenen Abschnitten behandelten Einzelmodellen, wobei diese über die Druckwellen miteinander gekoppelt sind (siehe folgender Abschnitt). Für die Berechnungen angehaltener Klänge stellte sich das CTIM-Verfahren als ungeeignet heraus, da der Algorithmus bei längeren Berechnungen im Fall divergenter Kegelsegmente instabil wurde. Mit dem Wellenleiter-Ansatz konnten sowohl Aspirationsrauschen als auch beliebige Vokalklänge nachgebildet werden.

Interaktion der Modellkomponenten

Die Simulationen mit der Kombination der Modelle für die Stimmlippen, das Ansatzrohr und die Abstrahlung lassen eine gegenseitige Abhängigkeit der Modelle erkennen. Dies lässt den Schluss zu, dass qualitativ hochwertige Stimmerzeugung nicht mit dem Oszillator-Filter-Modell realisierbar ist. Wurden bei der Kombination der Modelle die an den Grenzen zwischen den Modellen auftretenden Impedanz- und Schalldruckverläufe berücksichtigt, ließ sich eine deutliche Verbesserung der Natürlichkeit bei Vokalen erreichen.

Simulationsergebnisse

Das kombinierte Modell wurde zur Berechnung von stimmhaften Vokalklängen eingesetzt und, beispielhaft für die Berechnung eines komplexeren Klangs, für die Synthese von Obertonklängen verwendet. Weitere Analysemethoden (siehe Abschitt 8.3) wie die Sonographie eines Obertonsängers sowie analytische Berechnungen auf der Basis von Magnet-Resonanz-Verfahren bestätigten die These, dass die Überlagerung von zwei Resonanzen für die Bildung der Melodiestimme beim Obertonsingen verantwortlich ist. Die erste Resonanz wird durch einen $\lambda/2$-Längsresonator zwischen Glottis und einer Verengung im Ansatzrohr gebildet, die zweite von einem Helmholtzresonator zwischen der Verengung und der Mundöffnung.

Ein weiteres Beispiel der Anwendung des Modells ist die Nachbildung der Stimme eines Sängers mit Sängerknötchen. Sängerknötchen können bei starker Stimmbelastung auftreten und sind durch eine örtlich begrenzte Verdickung an der Kontaktfläche der Stimmlippen charakterisiert. Im Modell wurde eine solche Verdickung durch Vergrößerung der Masse des vierten von elf Segmenten nachgebildet. Die resultierende Schwingung wies einen unregelmäßigen Übergang zwischen der Grundmode und einer höheren Mode auf, wobei die verringerte Grundfrequenz durch Erhöhung der *vocalis*-Spannung wieder auf den Wert der Stimme ohne Sängerknötchen korrigiert werden konnte.

8.7 Schlussfolgerungen

Die vorliegende Arbeit beschreibt Methoden zur Berechnung und messtechnischen Erfassung von Schallsignalen im menschlichen Stimmapparat. Die Funktionskomponenten des Gesamtsystems „Sänger“ werden identifiziert und modelliert, wobei ein Ansatz im Zeitbereich gewählt wird, um die Zeitabhängigkeit und Interaktion der Funktionskomponenten berücksichtigen zu können.

Die Ergebnisse der Modellierung der Stimmlippenschwingung zeigen, dass das Mehrmassenmodell geeignet ist, auch komplexe Schwingungsvorgänge nachzubilden, obwohl eine sehr einfache Beschreibung der Jetablösung implementiert worden ist. Die Ergebnisse der Stimmsimulation werden mit Ergebnissen von Schalldruckmessungen an der Glottis verglichen und bestätigen eine geringe aber dennoch vorhandene Änderung des Stimmlippensignals in Abhängigkeit von der Ansatzrohrkonfiguration. Um die Ergebnisse der Simulationen zu überprüfen, wurden weitere *in situ*-Messungen zur Bestimmung der VTTF durchgeführt. Diese direkten Messungen setzen jedoch die recht invasive, transnasale Applikation eines Mikrofons in der Nähe der Stimmlippen voraus. Eine neu entwickelte Methode (VTMI) zur nicht-invasiven Bestimmung der Resonanzen des Ansatzrohrs wird vorgestellt, und Ergebnisse werden für verschiedene Stimmkonfigurationen gezeigt. Die Impedanzmethode bietet auch bei *in situ*-Messungen während der Phonation ein ausreichendes Signal-zu-Rausch-Verhältnis. Eine Anwendung der Methode für die Diagnose und Therapie von Artikulationsstörungen wird beschrieben. Die Verwendungsmöglichkeiten der Methode im klinischen Bereich sind Gegenstand aktueller Untersuchungen.

Appendix

Appendix A

Abbreviations – Terms – Symbols

Table A.1: List of abbreviations

Abbreviation	Description
Av.	Average
cf.	confer
CTIM	Continuous-time interpolated multiconvolution [BKC99]
dB	Decibel
DOF	Degree of freedom
EAF	Equivalent area function(s)
FEM	Finite element model
FFT	Fast Fourier transform
GNE	Glottal-to-noise excitation (ratio) [MGS97]
GUI	Graphical user interface
HNR	Harmonics-to-noise ratio
Hz	Hertz
IF	Ishizaka-Flanagan VF model [IF72]
IFFT	Inverse fast Fourier transform
ITA	Institute of Technical Acoustics, RWTH Aachen
KL	Kelly-Lochbaum VT model [KL62]
OQ	Open quotient of the glottal cycle
pl.	plural
Std.	Standard deviation
SNR	Signal-to-noise ratio
SPL	Sound pressure level
VF	Vocal fold(s)
vs.	*versus*: against
VT	Vocal tract
VTMI	Vocal tract impedance at the mouth
VTTF	Vocal tract transfer function

Table A.2: List of Latin terms

Latin term	**English description**	**German description**
ambitus	interval lowest ↔ highest note	Stimmumfang
anterior	towards the front	vordere(r)
caudal	towards the bottom	nach unten, untere(r)
cortex	cortex (periphery part of the brain)	Hirnrinde
diaphragma	diaphragm	Zwerchfell
epiglottis	epiglottis (area above the VF)	Kehldeckel
glottis (slit)	opening between the VF, glottis	Stimmritze
larynx	larynx (anatomical section near VF)	Kehlkopf
palatum	palate	harter Gaumen
peritoneum	abdominal movable tissue	Bauchfell
pharynx	throat	Rachen
posterior	towards the back	hintere(r)
velum	velum	weicher Gaumen
vestibulum oris	vestibule of the mouth	Mundvorhof
vortex	eddy	Wirbel
VOX	voice, program code	Stimme, Programmcode

Table A.3: List of symbols

Symbol	Description	Value	Unit
A	Area		m^2
a	Length of mass in y-direction		m
C_p	Thermal capacity at 0° C, 100 kPa	1000	$\mathrm{J\,kg^{-1}\,deg^{-1}}$
c_0	Speed of sound in air at 37° C, 100 % rel. humidity	350	$\mathrm{m\,s^{-1}}$
D	Damping constant		$\mathrm{kg\ s^{-1}}$
d	Diameter, thickness of VF in $x-$dimension		m
η	Nonlinearity coefficient, Adiabatic constant		$\mathrm{m^{-2}}$, -
F	Force		$\mathrm{kg\,m\,s^{-2}}$
f	Frequency		Hz
f_0	Fundamental frequency		Hz
g	Index for glottis		-
Γ	Vortex rotation		$\mathrm{m^2\,s^{-1}}$
H_1	1^{st} order Struve function		-
h	Pressure impulse response		Pa
K_1	Related Bessel function		-
J_0	1^{st} order Bessel function		-
k	Spring constant		$\mathrm{kg\,s^{-2}}$
l	Length		m
λ	Thermal conductivity at 0° C	$22.9 \cdot 10^{-3}$	$\mathrm{W\,m^{-1}\,deg^{-1}}$
L	Length of the vocal tract	0.14	m
M	Mach number		1
m	Index for *mucosa* mass or mouth		-
μ	Shear viscosity coefficient of air	$1.86 \cdot 10^{-5}$	$\mathrm{N\,s\,m^{-2}}$
ν	Kinetic viscosity of air	$1.5 \cdot 10^{-5}$	$\mathrm{m\,s^{-1}}$
P	Sound pressure		Pa
p	Sound pressure wave		Pa
R	Reflection coefficient		1
Re	Reynolds number		1
ρ_0	Air density at 37° C, 100 % relative humidity	1.14	$\mathrm{kg\,m^{-3}}$
S	Strain		1
St	Strouhal number		1
s	Signal		-
sub	Index for entity above the glottis		-
sup	Index for entity below the glottis		-
T	Transmission coefficient		1
th	Thickness of masses		m
U	Flow, Volume velocity		$\mathrm{m^3\,s^{-1}}$
$\dot{U}$	Flow derivative, Volume acceleration		$\mathrm{m^3\,s^{-2}}$
U_j	Jet velocity		$\mathrm{m\,s^{-1}}$
v	particle velocity, Index for *vocalis* mass		$\mathrm{m\,s^{-1}}$
ξ	Damping coefficient		1
Z_0	Acoustic impedance in air	414	$\mathrm{kg\,m^{-2}\,s^{-1}}$

Appendix B

Speech sounds

Table B.1: Speech sounds with examples in English and German

Sound	Description	English	German
[a]	unrounded open front vowel	—	f**a**sten
[ɑ]	unrounded open back vowel	f**a**r	V**a**ter
[ʌ]	unrounded mid-open back vowel	c**o**me	—
[e]	unrounded mid-closed front vowel	—	b**e**ten
[i]	unrounded closed front vowel	h**ee**d	M**ie**te
[ɪ]	unrounded half-closed front vowel	h**i**d	W**i**nd
[ɛ]	unrounded mid-open front vowel	h**ea**d	M**ä**nner
[æ]	unrounded half-open front vowel	h**a**d	**Ä**hre
[ɔ]	rounded mid-open back vowel	p**a**w	h**o**ffen
[o]	rounded mid-closed back vowel	—	B**oo**t
[u]	rounded closed back vowel	wh**o**	B**u**ch
[ʊ]	rounded half-closed back vowel	h**oo**d	M**u**tter
[y]	rounded closed front vowel	—	H**ü**te
[x]	velar voiceless fricative	—	a**ch**
[ʃ]	postalveolar voiceless fricative	**sh**ake	**Sch**nee
[s]	alveolar voiceless fricative	**s**ee	Spa**ß**
[j]	palatal voiceless fricative	—	**j**a
[ŋ]	velar nasal	si**n**ger	Säng**n**er

Table B.2: Range of formant frequencies for vowels (F1-F4), taken from [Val98] and poles in the spectra of the consonants.

Phoneme *Vowels*	1^{st} Formant [Hz]	2^{nd} Formant [Hz]	3^{rd} Formant [Hz]	4^{th} Formant [Hz]
[ɑ:]	610-850	1100-1900	3000-3100	3750-4100
[ɛ:]	290-650	1770-2300	2680-2870	3450-3880
[e:]	230-560	2300-2630	2800-3140	3600-3820
[i:]	160-460	2400-2500	3100-3400	3600-3700
[y:]	200-580	1660-2090	2310-2590	3320-3440
[o:]	230-480	640-920	2510-2710	-
[u:]	160-400	500-900	-	-
Consonants	1^{st} Pole	2^{nd} Pole	3^{rd} Pole	4^{th} Pole
[ʃ]	2200	4900	-	-
[x]	730-1100	3000-4000	-	-
[j]	260-440	2100-2700	2950-3450	3750-3950

Appendix C

Construction of the artificial singer

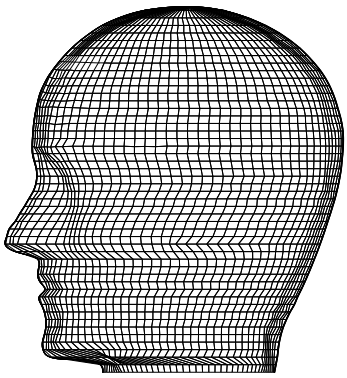


Figure C.1: Mesh of the artificial head

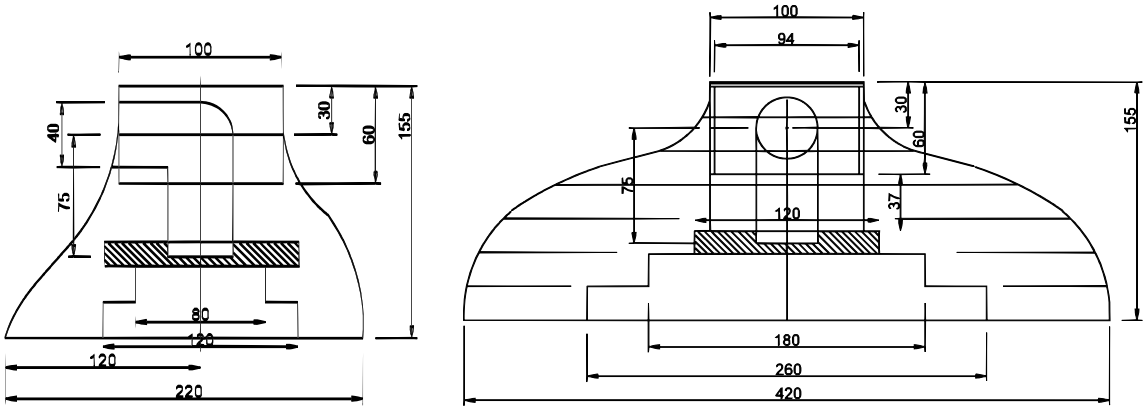


Figure C.2: View of torso from the side (left) and from the front (right), measures in mm

Appendix D

Forces on the *mucosa* masses

Tension, spring and contact forces

Similar to equation (2.23) in section 2.3.1, a tension force due to a polynomial stress-strain curve after [AT91] is assumed:

$$F_m^T = th_m d_m \left(0.5 + 38.3S_m - 49.5S_m^2 + 347.6S_m^3\right) \text{kPa} \; . \tag{D.1}$$

The spring forces on the *mucosa* masses in $x-$direction are similar to the *vocalis* mass forces but coupling force and contact force have opposite sign, and the boundary force does not exist:

$$\begin{aligned} F_{x,m,i}^T &= F_{m,i}^T \tfrac{x_{m,i-1}-x_{m,i}}{r_{m,i}} + F_{m,i+1}^T \tfrac{x_{m,i+1}-x_{m,i}}{r_{m,i+1}} \; , \\ F_{x,m,i}^v &= -F_{v,m,i}^m \; , \\ F_{x,m,i}^c &= \begin{cases} k_{x,m}^c \left(w_m + \eta_{k,x,m}^c w_m^3\right) & \text{for } w_m < 0 \; ; \\ 0 & \text{else.} \end{cases} \end{aligned} \tag{D.2}$$

The forces that act in $z-$direction on the *mucosa* masses can be described as follows:

$$\begin{aligned} F_{z,m,i}^T &= F_{m,i}^T \tfrac{z_{m,i-1}-z_{m,i}}{r_{m,i}} + F_{m,i+1}^T \tfrac{z_{m,i+1}-z_{m,i}}{r_{m,i+1}} \; , \\ F_{z,m,i}^v &= -F_{z,v,i}^m \; , \\ F_{z,m,i}^c &= -F_{z,v,i}^c \; . \end{aligned} \tag{D.3}$$

Damping forces

With the damping coefficients $D^m_{x,i}$ and $D^m_{z,i}$ that were given in the equations (2.36) for the connections of the *mucosa* masses to the *vocalis* masses, and the damping coefficient for the coupling springs to the neighboured *mucosa* elements

$$D^T_{m,i} = 2\xi^T_{m,i}\sqrt{m_m \frac{F^T_{m,i}}{r_{m,i}}}\ , \tag{D.4}$$

the damping force in $x-$direction on the *mucosa* mass is given to

$$F^d_{x,m,i} = D^T_{m,i}\dot{d}(x_{m,i} - x_{m,i-1}) - D^T_{m,i+1}\dot{d}(x_{m,i} - x_{m,i+1}) - D^m_{x,i}\dot{d}(x_{v,i} - x_{m,i})\ . \tag{D.5}$$

The damping force in $z-$direction is given to

$$F^d_{z,m,i} = -D^T_{m,i}\dot{d}(z_{m,i} - z_{m,i-1}) - D^T_{m,i+1}\dot{d}(z_{m,i} - z_{m,i+1}) - D^m_{z,i}\dot{d}(z_{v,i} - z_{m,i})\ . \tag{D.6}$$

Aerodynamic forces

The following cases must be distinguished for the *muscosa* mass in direction of the co-ordinate x:

$$F^a_{x,m} = \begin{cases} P_m th_m a & \text{for } A_m < A_v \text{ (convergent)}, \quad \text{open glottis;} \\ P_m th_m a & \text{for } A_m > A_v \text{ (divergent)}, \quad \text{open glottis;} \\ 0 & \text{for } A_m = 0, A_v > 0, \quad \text{closed glottis;} \\ P_m th_m a & \text{for } A_m > 0, A_v = 0, \quad \text{closed glottis;} \\ 0 & \text{for } A_m = A_v = 0, \quad \text{closed glottis.} \end{cases} \tag{D.7}$$

The pressure P_m on the *vocalis* mass has been given in (2.42).
For the *mucosa* mass, equation (D.8) gives the forces in the $z-$direction:

$$F^a_{z,m} = \begin{cases} P_v(w_v - w_m)a - P_{sup}d_m a & \text{for } A_m < A_v \text{ (convergent)}, \quad \text{open glottis;} \\ -P_{sup}d_m a & \text{for } A_m > A_v \text{ (divergent)}, \quad \text{open glottis;} \\ P_{sub}w_v a - P_{sup}d_m a & \text{for } A_m = 0, A_v > 0, \quad \text{closed glottis;} \\ -P_{sup}d_m a & \text{for } A_m > 0, A_v = 0, \quad \text{closed glottis;} \\ -P_{sup}d_m a & \text{for } A_m = A_v = 0, \quad \text{closed glottis.} \end{cases} \tag{D.8}$$

Appendix E

Program code of the waveguide model

```
data=get(Hndl.figureNumber, 'UserData');
radius = data.radius;
dist = data.spacing;
numcyl=length(radius); % number of cylinders constituting the VT model
lastj=numcyl+1; % index of last junction (=output end)
c0 = c; % sound speed in air [m/s]
Fs = round(c0/dist); % sampling frequency
dx=c0/Fs; % path increment [m]
dt=1/Fs; % time increment [s]
damp=1-Xi0;
nuem=zeros(1,numcyl-1); % reflection coefficient
nuep=zeros(1,numcyl-1); % refl. coeff. for backwards travelling waves
lcyld=ones(1,numcyl);
lcyl=lcyld*dist;
lsum=sum(lcyl); % total length of vocal tract
n=round(lsum/dx); % total number of samples
nspc=round(lcyl*n/lsum); % number of samplepoints within each cylinder
rl=radius;
rr=rl;
for i=2:numcyl
 Bm(i)=(rr(i-1)/rl(i))^2; % Sout/Sin
end
Bm(1)=1; % dummy value for division in next line
Bp=1./Bm;
a=-c0/(2*rr(numcyl));
a32=a; b3=a;
```

```
Teta32=[exp(b3*dt)]; % matrix Teta for output end
TT21=((b3*dt-1)+exp(-b3*dt))/(dt*b3^2);   % 2nd row, 1st col.
TT22=(1-(1+b3*dt)*exp(-b3*dt))/(dt*b3^2); % 2nd row, 2nd col.
TetaT32=[TT21 TT22]; % matrix TetaTilde for output end
A32=Teta32*TetaT32; % matrix A (1 row, 2 columns) for output end
rvoc = radius(1);
vocarea = pi*(rvoc^2);

% area ratio at node glottis->VT (outgoing wave)
Bglotm  = glotarea/vocarea;
if glotarea == 0
  Bglotp = 1e10;
else
  % area ratio at node VT->glottis (ingoing wave)
  Bglotp  = 1/Bglotm;
end;
T1m = 2*glotarea/(vocarea+glotarea);

% start of calculus routine
pmr2.junction(1).time(1) = ppR.junction(1).time(1) + out_vf_voc*T1m;

% reflection and transmission of outgoing wave at input end (glottis)
pmR.junction(1).time(1) = out_vf_voc*(Bglotm-1)/(Bglotm+1);
pmT.junction(1).time(1) = pmR.junction(1).time(1) + out_vf_voc;

% add noise
pmT.junction(1).time(1) = pmT.junction(1).time(1) + PARNOISE.out(1);

% reflection and transmission of outgoing wave
for i=2:numcyl
  pmR.junction(i).time(1)=pmr2.junction(i-1).time(1)*...
    (Bm(i)-1)/(Bm(i)+1);          % reflected part of pmr2
  pmT.junction(i).time(1)=pmR.junction(i).time(1)+...
    pmr2.junction(i-1).time(1);   % transmitted part of pmr2

% add noise
pmT.junction(i).time(1)= pmT.junction(i).time(1) + PARNOISE.out(i);
end
```

```
% reflection of outgoing wave at output end;
if radimproutine == 1 % simple radiation impedance
zmR.junction(lastj).time(1) = ...
zmR.junction(lastj).time(2)*Teta32 ...
+ A32(1)*pmr2.junction(numcyl).time(2) ...
+ A32(2)*pmr2.junction(numcyl).time(1);
pmR.junction(lastj).time(1) = a32*zmR.junction(lastj).time(1);
p_out_mouth = ...
pmR.junction(lastj).time(1)...
+ pmr2.junction(lastj-1).time(1);
elseif radimproutine == 2  % digital filter
  Rm.state(:,2) ...
= Rm.A*Rm.state(:,1) ...
+ Rm.B*pmr2.junction(numcyl).time(1);
  pmR.junction(lastj).time(1) ...
= Rm.C*Rm.state(:,1) ...
+ Rm.D*pmr2.junction(numcyl).time(1);
Rm.state(:,1) = Rm.state(:,2);
p_out_mouth ...
= pmR.junction(lastj).time(1)...
+ pmr2.junction(lastj-1).time(1);
end

% reflection and transmission of ingoing wave
for i=numcyl:-1:2
  ppR.junction(i).time(1)=ppl2.junction(i+1).time(1)*...
    (Bp(i)-1)/(Bp(i)+1);          % reflected part of ppl2
  ppT.junction(i).time(1)=ppR.junction(i).time(1)+...
    ppl2.junction(i+1).time(1);   % transmitted part of ppl2
end

% reflection and transmission of ingoing wave at input end (glottis)
ppR.junction(1).time(1) = ppl2.junction(2).time(1);

% timeshift of outgoing and ingoing waves
for i=1:numcyl
  pmr2.junction(i).time(2:lastp2(i))=...
```

```
    pmr2.junction(i).time(1:lastp2(i)-1);
  ppl2.junction(i+1).time(2:lastp2(i))=...
    ppl2.junction(i+1).time(1:lastp2(i)-1);
end
zmR.junction(lastj).time(2)=zmR.junction(lastj).time(1);
% end of timeshift

% newest element of outgoing wave
for i=2:numcyl
  pmr2.junction(i).time(1)=ppR.junction(i).time(1)+...
    pmT.junction(i).time(1);
  pmr2.junction(i).time(1)=damp*pmr2.junction(i).time(1);
end
pmr2.junction(1).time(1)=damp*ppR.junction(1).time(1);

%newest element of ingoing wave
ppl2.junction(lastj).time(1)=pmR.junction(lastj).time(1);
for i=numcyl:-1:2
  ppl2.junction(i).time(1)=pmR.junction(i).time(1)+...
    ppT.junction(i).time(1);
  ppl2.junction(i).time(1)=damp*ppl2.junction(i).time(1);
end

%newest element of ingoing wave into vocal folds
pin_gl_ac = pmR.junction(1).time(1)+ppT.junction(1).time(1);
```

Appendix F

Program code of the multiconvolution model

The MATLAB code is represented as pseudo code.

1. define physical parameters (c_0, ρ_0, ξ, vocal tract dimensions)
2. calculate constants
   $(degree, b, \alpha_1, \alpha_2, ldis, m_{inf}, m_{sup}, C_+, C_-, A_{input}, B_{input}, C_{input}, A, B)$
   - $coefsigma = \frac{\sqrt{2/\pi}\xi L}{c_0^{1.5}(d1+d2)}$
   - $expsigma = \frac{2}{c_0^3}(\frac{\xi L}{d1+d2})^2$
   - $t_0 = (\frac{L}{c_0 \Delta t} - floor(\frac{L}{c_0 \Delta t}))\Delta t$
   - $\sigma = coefsigma(t - t_0)^{-1.5} e^{(\frac{-expsigma}{t-t0})}$
   - $k[m, s] = \int_{t_0}^{\Delta t} \sigma \cdot (\frac{t}{\Delta t})^s dt$
3. iterative calculation of the pressure waves at discontinuities (1..ldis):
   - **Application of damping α to the pressure waves $p_+, p_-, output_+, output_-$ at all discontinuities:**
     $p_+(t) = input_+/b$
     $p_+(t + \Delta t) = p_+(t)$ % rotateright
     $p_+(t) = 2p_+(t + \Delta t) - p_+(t + 2\Delta t)$
     $p_-(t) = input_- \cdot b$
     $p_-(t + \Delta t) = p_-(t)$ % rotateright
     $p_-(t) = 2p_-(t + \Delta t) - p_-(t + 2\Delta t)$
     $output_+(t + \Delta t) = output_+(t)$ % rotateright
     $output_+(t) = \alpha_1 \bullet p_+(m_{inf} \,..\, m_{sup}) + \alpha_2 \bullet p_+(m_{inf} + 1 \,..\, m_{sup} + 1)$ [1]

[1] $\bullet$ symbolizes the dot product

$output\_(t + \Delta t) = output\_(t)$ (rotateright)

$output\_(t) = \alpha_1 \bullet p\_(m_{inf} \,..\, m_{sup}) + \alpha_2 \bullet p\_(m_{inf} + 1 \,..\, m_{sup} + 1)$

- **Set actual flow to $\delta(t)$ (Dirac function):**
- **Application of matrices to entry discontinuity $i = 1$ ($input\_, flowinput, y_+aux, inputaux$):**

  $flowinput(t + \Delta t) = flowinput(t)$ % rotateright

  $flowinput(1) =$ actual flow

  $y_+aux(1) = A(1) \bullet y_+aux(1) + B(1) \bullet output_+(1)$

  $inputaux = A_{input} \bullet inputaux + B_{input} \bullet flowinput$

  $input\_(1) = C_+(1) \bullet y_+aux(1) + C_{input} \bullet inputaux$
- **Application of matrices to middle discontinuities i ($input_+, input\_, y_+aux, y\_aux$):**

  $y_+aux(i) = A(i) \bullet y_+aux(i) + B(i) \bullet output_+(i)$

  $y\_aux(i) = A(i) \bullet y\_aux(i) + B(i) \bullet output\_(i - 1)$

  $aux = C_+(i) \bullet y_+aux(i) + C\_(i) \bullet y\_aux(i)$

  $input_+(i - 1) = aux + output_+(i, t)$

  $input\_(i) = aux + output\_(i - 1, t)$
- **Application of matrices to exit discontinuity $i = ldis$ ($input_+, y\_aux$):**

  $y\_aux(ldis) = A(ldis) \bullet y\_aux(ldis) + B(ldis) \bullet output\_(ldis - 1)$

  $input_+(ldis - 1) = C\_(ldis) \bullet y\_aux(ldis)$
- **Addition of inward and outward wave yield the pressure at discontinuity $ndis$:**

  $p(t) = output_+(ndis) + input\_(ndis)$

Appendix G

Graphical user interfaces

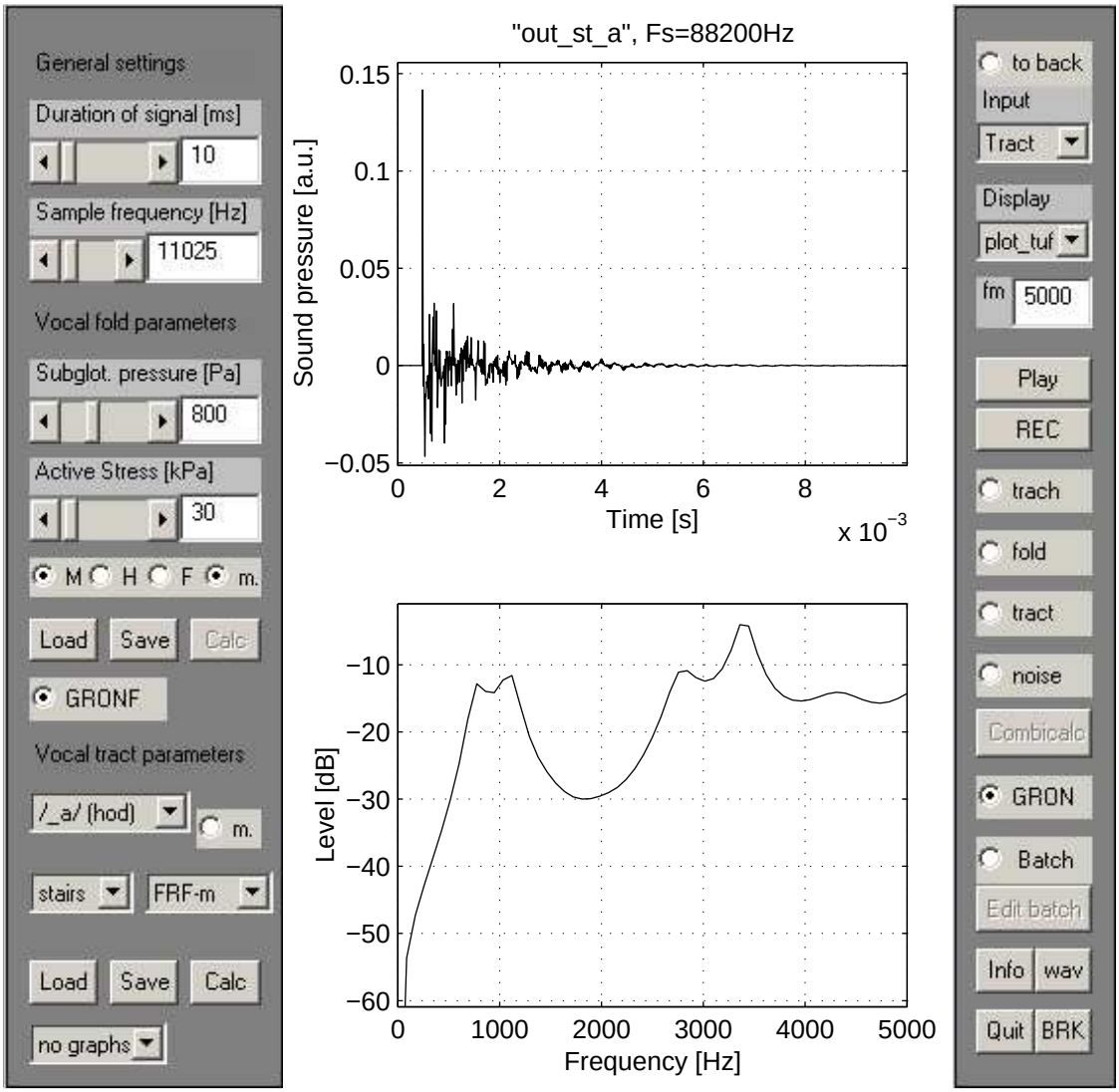


Figure G.1: Main window of VOX

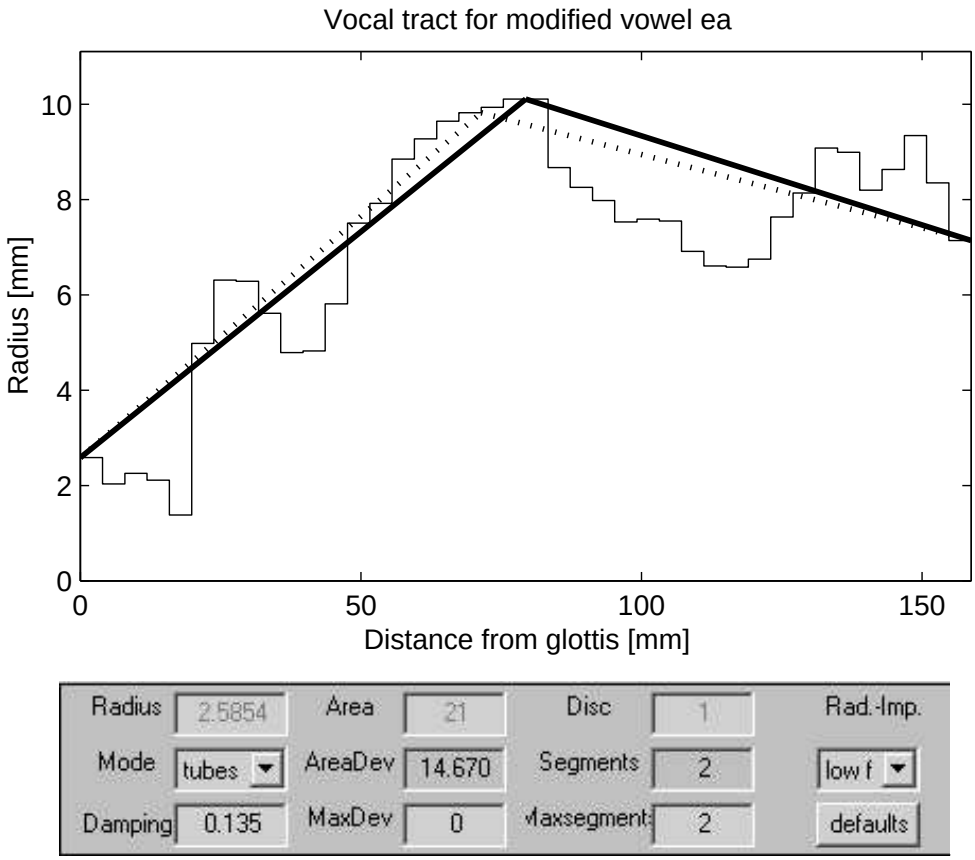


Figure G.2: Interface for design of the EAF of the VT

List of Tables

List of Figures

Acknowledgements

The author wishes to thank Professor Dr. rer. nat. Michael Vorländer for supervision of the thesis and for the opportunity to explore the fascinating world of human sound generation. I am grateful to Professor Dr.-Ing. Peter Vary and Professor Dr.-Ing. Jürgen Meyer for their interest in the subject and their friendly willingness of being correctors of this thesis.

Harald Jers, Ulrich Reiter, Nils Alhäuser, Daniel Riemann, Philipp Sellerbeck and Philipp Heck (in chronological order) contributed with their work at the Institute of Technical Acoustics to this thesis. The electroacoustic group at ITA, especially Gottfried Behler and Jochen Kleber, contributed significantly with their knowledge to the construction and equalisation of the artificial singer. However, the realisation of the singer and of the measurement devices would have been impossible without the great work of the ITA workshop people Franz Buchholz, Hans-Jürgen Dilli, Rolf Kaldenbach and Uwe Schlömer.

Mico Hirschberg et Xavier Pelorson ont evocés mon intérêt pour le sujet. Merci pour notres discussions sur l'aerodynamique entre les cordes vocales. Merci beaucoup, Pierre Badin et Coriandre Vilain, pour notres discussions pendant mon séjour au ICP à Grenoble. Un grand merci à Ana Barjau pour le CTIM algorithme.

Wolfgang Saus, the singers from 'Huun Huur Tu' and many students who gave their voice for measurements are acknowledged for their patience. Prof. Dr. med. Christiane Neuschaefer-Rube is acknowledged for assistance during invasive measurements of the vocal tract transfer functions.

I would like to thank Sonja Meyer zu Berstenhorst for the wonderful moments of my social life besides university and for the support during the last months, days and minutes of writing this thesis.

Bernd J. Kröger is thanked for proof-reading the manuscript and for stimulating discussions on notation of phonemes and speech generation – and thanks for the vector! Tatjana von Stackelberg gave advice regarding the medical terms and Calum Gray did a great job reading through the whole manuscript and improving my German English.

Finally, a very special thanks goes to my parents and the rest of the family for the confidence in the success of this thesis.

Malte Kob Aachen, January 31st 2002

Curriculum vitae

15.12.1967	born in Hamburg, Germany

Education

1974-1978	Primary school, Tornesch-Esingen
1978-1988	Secondary school "Elsa-Brändström-Schule", Elmshorn
1988-1989	Military service, Pinneberg
1987-1989	C-Kurs (Lessons in church music), Hamburg

Course of studies

1989-1994	Studies of electrical engineering at the Technical University Braunschweig
1994	Diplomarbeit at the Physikalisch-Technische Bundesanstalt (PTB), Braunschweig
1996-2002	Ph. D. at the Aachen University (RWTH)

Employments

1993-1994	Practical training at Deutsche Grammophon, Hannover
1994-1995	Scientific employee at the PTB Braunschweig
1996-2001	Scientific employee at the Institute of Technical Acoustics (ITA), RWTH
since 2001	Scientific employee at the Department of Phoniatrics, Pedaudiology and Communication Disorders, University Hospital Aachen, Faculty of medicine, RWTH

Bibliography

[AB01] F. Alipour and D. A. Berry, *Effects of glottal asymmetry on modes of phonation*, File 8_16.pdf on CD-ROM IV of the 17^{th} ICA (Rome), 17^{th} International Congress on Acoustics, 2001, pp. 6–8.

[Alh99] N. Alhäuser, *Ein physikalisches Modell für die menschlichen Stimmlippen*, Diplomarbeit, Aachen University, Institute of Technical Acoustics, 1999.

[AT91] F. Alipour-Haghighi and I. R. Titze, *Elastic models of vocal fold tissues*, J. Acoust. Soc. Am. **90** (1991), no. 3, 1326–1331.

[AT96] F. Alipour and I. Titze, *Combined simulation of two-dimensional airflow and vocal fold vibration*, Vocal Fold Physiology: Controlling complexity and chaos (P. Davis and N. Fletcher, eds.), Singular Pub. Group, San Diego, 1996, pp. 17–29.

[ATP89] F. Alipour-Haghighi, I. Titze and A. Perlman, *Tetanic contraction on vocal fold muscle*, J. Speech Hera. Res. **32** (1989), 226–231.

[AY99] S. Adachi and M. Yamada, *An acoustical study of sound production in biphonic singing, Xöömij*, J. Acoust. Soc. Am. **105** (1999), no. 5, 2920–2932.

[BF94] P. Biesalski and F. Frank, *Phoniatrie – Pädaudiologie*, Georg Thieme, Stuttgart, New York, 1994.

[BKC99] A. Barjau, D. H. Keefe and S. Cardona, *Time-domain simulation of acoustical waveguides with arbitrarily spaced discontinuities*, J. Acoust. Soc. Am. **105** (1999), no. 3, 1951–1964.

[BMB+96] P. Badin, K. Mawass, G. Bailly, C. Vescovi, D. Beautemps and X. Pelorson, *Articulatory synthesis of fricative consonants: data and models*, Proceddings of the Fourth Speech Production Seminar – First ESCA Tutorial and Research Workshop on Speech Production Modeling: from Control Strategies to Acoustics (Autrans, France), 1996, pp. 221–224.

[Boe93] H. G. Boenninghaus, *Hals-Nasen-Ohrenheilkunde für Medizinstudenten*, Springer, Berlin, Heidelberg, New York, 1993.

[CHM+90] D. Childers, D. Hicks, G. Moore, L. Eskenazi and A. Lalwani, *Electroglottography and vocal fold physiology*, J. Speech Res. **33** (1990), 245–254.

[CHMA86] D. Childers, D. Hicks, G. Moore and Y. Alsaka, *A model for vocal fold vibratory motion, contact area, and the electroglottogram*, J. Acoust. Soc. Am. **80** (1986), 1309–1320.

[Coo90] P. R. Cook, *Identification of control parameters in an articulatory vocal tract model, with applications to the synthesis of singing*, Ph. D. thesis, Stanford University, 1990.

[DA01] C. Drioli and F. Avanzini, *A physically-informed model of the glottis with application to voice quality assessment*, Proceedings of the 2^{nd} International workshop on Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA) (Florence, Italy) (C. Manfredi, ed.), University of Florence, 2001.

[dB02] H. E. de Bree, *An overview of microflown technologies*, Acustica – acta acustica **88** (2002), no. 3,
further information at URL `http://www.microflown.com`.

[Dre01] C. Dresel, *Biomechanical modeling of the human voice: A modified two-mass model*, Diplomarbeit, Friedrich-Alexander-Universität Erlangen-Nürnberg, 2001.

[DSW96] A. Dowd, J. Smith and J. Wolfe, *Real Time, Non-Invasive Measurement of Vocal Tract Resonances: Application to Speech Training*, Acoustics Australia **24** (1996), no. 2, 53–60.

[DSW97] A. Dowd, J. Smith and J. Wolfe, *Learning to Pronounce Vowel Sounds in a Foreign Language using Acoustic Measurements of the Vocal Tract as Feedback in Real Time*, Language and Speech **41** (1997), no. 1, 1–20.

[Dud70] D. Dudgeon, *Two-mass model of the vocal cords*, J. Acoust. Soc. Am. **45** (1970), 118A.

[EPSB98] S. El-Masri, X. Pelorson, P. Saguet and P. Badin, *Development of the transmission line matrix method in acoustics applications to higher modes in the vocal tract and other complex ducts*, Int. J. Numer. Model. **11** (1998), 133–151.

[ESW97] J. Epps, J. R. Smith and J. Wolfe, *A novel instrument to measure acoustic resonances of the vocal tract during phonation*, Meas. Sci. Technol. **8** (1997), 1112–1121.

[FC69] L. J. Flanagan and L. Cherry, *Excitation of vocal-tract synthesizers*, J. Acoust. Soc. Am. **45** (1969), 764–769.

[FHKL97] G. Fant, S. Hertegård, A. Kruckenberg and J. Liljencrants, *Covariation of subglottal pressure, F0 and glottal parameters*, Proc. Eurospeech '97 (Rhodos), 1997, pp. 453–456.

[Fis93] P. M. Fischer, *Die Stimme des Sängers*, Metzler Stuttgart, 1993.

[FL71] O. Fujimura and J. Lindquist, *Sweep-tone measurements of vocal-tract characteristics*, J. Acoust. Soc. Am. **49** (1971), 541–558.

[Fla60] J. L. Flanagan, *Analog measurements of sound radiation from the mouth*, J. Acoust. Soc. Am. **32** (1960), 1613–1620.

[Fla65] J. L. Flanagan, *Speech Analysis Synthesis and Perception*, Springer, Berlin, Heidelberg, New York, 1965.

[Fuk99] L. Fuks, *From Air to Music – Acoustical, Physiological and Perceptual Aspects of Reed Wind Instrument Playing and Vocal-Ventricular Fold Phonation*, Ph. D. thesis, Royal Institute of Technology, Stockholm, Sweden, 1999.

[Goo96] M. Goodwin, *Residual modeling in music analysis-synthesis*, Proc. ICASSP, IEEE, 1996, pp. 1005–1008.

[Hai00] T. Q. Hai, *New Experiments on Overtone Singing*, Dokumentation 3. Internationale Stuttgarter Stimmtage 2000, Akademie für Gesprochenes Wort, Staatliche Hochschule für Musik und Darstellende Kunst Stuttgart, 2000, pp. 13–14.

[Hai01] T. Q. Hai, *Overtones – Recherches introspectives sur le chant diphonique et leurs applications*, 2001,
URL: `http://mapage.noos.fr/chechnik/diphointros.htm`.

[HDS+01] U. Hoppe, M. Döllinger, S. Schuberth, F. Rosanowski and U. Eysholdt, *Clinical significance of the two-mass-model*, Proceedings of the 2^{nd} International workshop on Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA) (Firenze, Italy) (A. Manfredi, ed.), University of Florence, 2001.

[Her91] D. J. Hermes, *Synthesis of breathy vowels: Some research methods*, Speech Communication **10** (1991), 497–502.

[HG80] T. Q. Hai and D. Guillou, *Original research and acoustical analysis in connection with the Xöömij style of biphonic singing*, pp. 162–173, Heibonsha, Tokyo, 1980.

[HG93] S. Hertegård and J. Gauffin, *Voice source-vocal tract interaction during high-pitched female singing*, Proceedings of the Stockholm music acoustics conference (SMAC) (Stockholm) (E. J. A. Friberg, J. Iwarsson and J. Sundberg, eds.), Royal Swedish Academy of Music, 1993, pp. 177–181.

[Hir68] M. Hirano, *The vocal cord during phonation*, Igaku no Ayumi **80** (1968), no. 10.

[Hir92] A. Hirschberg, *Some fluid dynamic aspects of speech*, Bulletin de la Communication Parlée **2** (1992), 7–30.

[HKW95] A. Hirschberg, J. Kergomard and G. Weinreich (eds.), *Mechanics of musical instruments*, ch. Aero-acoustics of wind instruments, pp. 291–369, Springer, 1995.

[HMOI92] T. Haji, K. Mori, K. Omori and N. Isshiki, *Mechanical Properties of the Vocal Fold*, Acta Otolaryngol **112** (1992), 559–565.

[Hof98] G. C. J. Hofmans, *Vortex Sound in Confined Flows*, Ph. D. thesis, Technische Universiteit Eindhoven, 1998.

[Hol74] H. Hollien, *On vocal registers*, Journal of Phonetics **2** (1974), 125–143.

[HPH+96] A. Hirschberg, X. Pelorson, G. C. J. Hofmans, R. R. van Hassel and A. P. J. Wijnands, *Starting Transient of the Flow Through an In-Vitro Model of the Vocal Folds*, Vocal Fold Physiology Conference: Controlling Complexity and Chaos (1996), 34–44.

[Hus62] R. Husson, *Physiologie de la phonation*, Paris, 1962.

[IF72] K. Ishizaka and J. L. Flanagan, *Synthesis of voiced sounds from a two-mass model of the vocal cords*, Bell Syst. Tech. J. **50** (1972), 1233–1268.

[IK68] K. Ishizaka and T. Kaneko, *On Equivalent Mechanical Constants of the Vocal Cords*, J. Acoust. Soc. Japan **24** (1968), no. 5, 312–313.

[IM72] K. Ishizaka and M. Matsudaira, *Fluid mechanical considerations of vocal cord vibration*, Tech. Report 8, Speech Commun. Res. Lab., Santa Barbara, CA, 1972.

[Jac00] P. J. Jackson, *Characterisation of plosive, fricative and aspiration components in speech production*, Ph. D. thesis, University of Southampton, 2000.

[Jer98] H. Jers, *Untersuchung der Realisierungsmöglichkeiten verteilter Quellen für die raumakustische Computersimulation am Beispiel des Chores*, Diplomarbeit, Aachen University, Institute of Technical Acoustics, 1998.

[JK98] H. Jers and M. Kob, *Nachbildung eines Chores für raumakustische und musikalische Untersuchungen*, Bericht zur 20. Tonmeistertagung Karlsruhe (Saur München), Verein Deutscher Tonmeister (VDT), 1998, pp. 208–217.

[Jov98] S. T. Jovičić, *Formant feature differences between whispered and voiced sustained vowels*, Acustica – acta acustica **84** (1998), 739–743.

[JS98] P. J. B. Jackson and C. H. Shadle, *Pitch-synchronous Decomposition of Mixed-source Speech Signals*, Proc. Joint Int. Cong. on Acoust. and Acoust. Soc. Am. (Seattle, WA 1), vol. 1, ICA, 1998, pp. 263–264.

[JS00] P. J. B. Jackson and C. Shadle, *Aero-acoustic modeling of voiced and unvoiced fricatives based on MRI data*, Proc. SPS5 (Seeon, Germany), 2000, pp. 185–188.

[JZ01] J. J. Jiang and Y. Zhang, *Modeling of chaotic vibrations in symmetric vocal folds*, J. Acoust. Soc. Am. **110** (2001), 2120–2128.

[KJ99] M. Kob and H. Jers, *Directivity measurement of a singer*, CD-ROM with the collected papers from the joint meeting Forum Acusticum/ASA/DAGA in Berlin, Acoustical Society of America, 1999, p. 2AMU_19.

[KL62] J. L. Kelly and C. C. Lochbaum, *Speech synthesis*, Proceedings of the 4th International Congress on Acoustics, 1962, Reprinted in: J. L. Flanagan and L. R. Rabiner (Editors): Speech Synthesis (Dowden, Hutchinson & Ross, Stoudsburg), S. 127-130, pp. 1–4.

[Kle96] J. Kleber, *Aufbau eines erweiterten digitalen Controllers für Lautsprecher (HUGO)*, Diplomarbeit, Aachen University, Institute of Technical Acoustics, 1996.

[Kle99] N. H. Kleinsasser, *The Kleinsasser's archive of microlaryngoscopy*, 1999, CD-ROM, ISBN 3-89756-039-9.

[Kli86] F. Klingholz, *Die Akustik der gestörten Stimme*, Thieme Verlag, Stuttgart, New York, 1986.

[KN02] M. Kob and C. Neuschaefer-Rube, *A method for measurement of the vocal tract impedance at the mouth*, Accepted for publication in Medical Engineering and Physics, 2002.

[KRJ+95] J. A. Koufman, T. A. Radomski, G. M. Joharji, G. B. Russell and D. C. Pillsbury, *Laryngeal Biomechanics Of The Singing Voice*, 1995, Presented at the Annual Meeting of the American Academy of Otolaryngology – Head and Neck Surgery, New Orleans, Louisiana.

[Kro95] W. Kropp, *Beschreibung akustischer Vorgänge im Zeitbereich*, unpublished postdoctoral thesis, 1995.

[Krö98] B. J. Kröger, *Ein phonetisches Modell der Sprachproduktion*, Niemeyer, Tübingen, 1998.

[Krö00] B. J. Kröger, *Analyse von MRT-Daten zur Entwicklung eines vokalischen Artikulationsmodells auf der Ebene der Areafunktion*, Elektronische Sprachsignalverarbeitung (Dresden) (K. Fellbaum, ed.), w.e.b. Universitätsverlag, 2000, pp. 201–208.

[KTH87] T. Koizumi, S. Taniguchi and S. Hiromitsu, *Two-mass models of the vocal cords for natural sounding voice synthesis*, J. Acoust. Soc. Am. **82** (1987), no. 4, 1197–1192.

[KWMP00] B. J. Kröger, R. Winkler, C. Mooshammer and B. Pompino-Marschall, *Estimation of vcal tract area function from magnetic resonance imaging: Preliminary results*, Proceedings of the 5^{th} Seminar on Speech Production: Models and Data (Kloster Seeon, Bavaria), 2000, pp. 333–336.

[LHVH98] N. J. C. Lous, G. C. J. Hofmans, R. N. J. Veldhuis and A. Hirschberg, *A Symmetrical Two-Mass Vocal-Fold Model Coupled to Vocal Tract and Trachea, with Application to Prosthesis Design*, Acustica – acta acustica **84** (1998), 1135–1150.

[Lil85] J. Liljencrants, *Speech synthesis with a reflection-type line analog*, Ph. D. thesis, Royal Institute of Technology, Stockholm, Sweden, 1985.

[Lil89] J. Liljencrants, *Numerical simulation of glottal flow*, Proc. Congr. of Vocal Fold Physiology (Stockholm), 1989.

[Lil91] J. Liljencrants, *A translating and rotating mass model of the vocal folds*, STL-QPSR **1** (1991), 1–18.

[LL98] M. Liu and A. Lacroix, *Analysis of the vocal tract for fricatives based on the acoustic tube model*, Fortschritte der Akustik – DAGA 98 (Oldenburg, Germany) (D. A. Sill, ed.), Deutsche Gesellschaft für Akustik (DEGA) e.V., 1998, pp. 346–347.

[Luc96] J. C. Lucero, *Chest and falsetto-like oscillations in a two-mass model of the vocal folds*, J. Acoust. Soc. Am. **100** (1996), 3355–3359.

[MA88] J. Martínez and J. Agulló, *Conical bores. Part I: Reflection functions associated with discontinuities*, J. Acoust. Soc. Am. **84** (1988), no. 5, 1613–1619.

[MAC88] J. Martínez, J. Agulló and S. Cardona, *Conical bores. Part II: Multiconvolution*, J. Acoust. Soc. Am. **84** (1988), no. 5, 1620–1627.

[Mae77] S. Maeda, *On a simulation method of dynamically varying vocal tract: reconsideration of the Kelly-Lochbaum model*, pp. 281–288, GALF, 1977.

[Mer98] P. Mergell, *Nonlinear dynamics of phonation – high-speed glottography and biomechanical modeling of vocal fold oscillations*, Ph. D. thesis, Technical University Berlin, 1998.

[MEV78] R. Monsen, A. Engebretson and N. Vemula, *Indirect assessment of the contribution of subglottal air pressure and vocal-fold tension to changes of fundamental frequency in English*, J. Acoust. Soc. Am. **64** (1978), 65–80.

[MGS97] D. Michaelis, T. Gramß and H. Strube, *Glottal to noise excitation ratio – a new measure for describing pathological voices*, Acustica – acta acustica **83** (1997), 700–706.

[MM84] J. Meyer and A. H. Marshall, *Schallabstrahlung und Gehörseindruck beim Sänger*, 13. Tonmeistertagungsbericht (Saur München), Verein Deutscher Tonmeister (VDT), 1984, pp. 232–336.

[MM95] E. Mommertz and S. Müller, *Measuring impulse responses with digitally pre-emphasized pseudorandom noise derived from maximum-length sequences*, Applied Acoustics **44** (1995), 195–214.

[MM01] S. Müller and P. Massarani, *Transfer Function Measurement with Sweeps*, J. Audio Eng. Soc. **49** (2001), no. 6, 443–471.

[Mül37] J. Müller, *Handbuch der Physiologie der Menschen*, Holscher, Koblenz, 1837.

[Ned69] C. Nederveen, *Acoustical aspects of woodwind instruments*, Ph. D. thesis, Frits Knut, Amsterdam, 1969.

[Net99] F. H. Netter, *Interaktiver Atlas der Anatomie des Menschen*, 1999, CD-ROM with Illustrations of the human body.

[NR00] C. Neuschaefer-Rube, *Entwicklung und Erprobung eines multiparametrischen Konzeptes zur Funktionsbeurteilung der Stimme*, unpublished postdoctoral thesis, Aachen University, 2000.

[NR01] C. Neuschaefer-Rube, *Bewegung des Ansatzrohres beim Singen*, personal communication, 2001.

[P+94] X. Pelorson *et al.* , *Theoretical and experimental study of quasisteady-flow separation within the glottis during phonation. Application to a modified two-mass model*, J. Acoust. Soc. Am. **96** (1994), no. 6, 3416–3431.

[Pha95] Y. Pham Thi Ngoc, *Caractérisation Acoustique Du Conduit Vocal: Fonctions de Transfert Acoustiques et Sources de Bruit*, Ph. D. thesis, Institut National Polytechnique de Grenoble, 1995.

[PHWB95] X. Pelorson, A. Hirschberg, A. Wijnands and H. Bailliet, *Description of the flow through in-vitro models of the glottis during phonation*, acta acustica **3** (1995), 191–202.

[PLK95] X. Pelorson, J. Liljencrants and B. Kröger, *On the aeroacoustics of voiced sound production*, Proc. International Congress on Acoustics (Trondheim), vol. IV, 1995, pp. 501–504.

[Rei99] U. Reiter, *Parameter eines Röhrenmodells für den menschlichen Vokaltrakt*, Diplomarbeit, Aachen University, Institute of Technical Acoustics, 1999.

[Rio01] V. Rioux, *Analysis of flue organ pipes - An interdisciplinary study*, Ph. D. thesis, Chalmers University of Technology, 2001.

[Rod95] X. Rodet, *One and two mass model oscillations for voice and instruments*, ICMC 95 (Centre Georges-Pompidou, Paris), IRCAM, 1995, URL: `http://mediatheque.ircam.fr/articles/textes/Rodet95/`.

[Sch56] R. Schilling, *Das kindliche Sprechvermögen*, Freiburg, 1956.

[Sch66] W. Schilz, *Richtcharakteristik der Schallabstrahlung einer durchströmten Öffnung*, Acustica **17** (1966), 364–366.

[Sch95] A. Schmitz, *A new digital artificial head measuring system*, Acustica **81** (1995), no. 4, 416–420.

[Sch99] M. R. Schroeder, *Computer Speech – Recognition, Compression, Synthesis*, Springer-Verlag, Berlin, Heidelberg, New York, 1999.

[Sch01] J. Schoentgen, *Stochastic models of jitter*, J. Acoust. Soc. Am. **109** (2001), 1631–1650.

[Sel01] P. Sellerbeck, *Ein physikalisches Modell zur Synthese des Rauschanteils der menschlichen Singstimme*, 2001, Studienarbeit, Institute of Technical Acoustics.

[Sen00] Sennheiser electronic KG, *KE 4 – Industry information*, D-30900 Wedemark, 2000.

[SG71] M. M. Sondhi and B. Gopinath, *Determination of vocal-tract shape from impulse response at the lips*, J. Acoust. Soc. Am. **49** (1971), 1867–1873.

[SH95] I. Steinecke and H. Herzel, *Bifurcations in an asymmetric vocal fold model*, J. Acoust. Soc. Am. **97** (1995), 1571–1578.

[SH00] H. Saweda and S. Hashimoto, *Mechanical construction of a human vocal system for singing voice production*, Advanced Robotics **13** (2000), no. 7, 647–661.

[Sha85] C. Shadle, *The acoustics of fricative consonants*, Ph. D. thesis, MIT, Cambridge, MA, 1985.

[Sin99] D. J. Sinder, *Speech Synthesis Using an Aeroacoustic Fricative Model*, Ph. D. thesis, University of New Jersey, 1999.

[SK00] M. Sapp and J. Kleber, *Universal Audio Signal Processing System*, Acustica – acta acustica **86** (2000), 185–186.

[Smi96] J. O. Smith, *Physical Modeling Synthesis Update*, Computer Music Journal **20** (1996), no. 2, 44–56.

[SS87] M. M. Sondhi and J. Schröter, *A Hybrid Time-Frequency Domain Articulatory Speech Synthesizer*, IEEE Trans. on Acoustics, Speech, and Signal Processing **ASSP-35** (1987), no. 7, 955–967.

[SSH+81] B. C. Sonies, T. H. Shawker, T. E. Hall, L. H. Gerber and S. B. Leighton, *Ultrasonic visualization of tongue motion during speech*, J. Acoust. Soc. Am. **70** (1981), 683–686.

[ST95] B. H. Story and I. R. Titze, *Voice simulation with a body-cover model of the vocal folds*, J. Acoust. Soc. Am. **97** (1995), no. 2, 1249–1260.

[ST96] B. H. Story and I. R. Titze, *Vocal tract area functions from magnetic resonance imaging*, J. Acoust. Soc. Am. **100** (1996), no. 1, 537–554, Area functions used in vox.

[Str82] H. Strube, *Time-varying wave digital filters and vocal tract models*, IEEE Proc. ICASSP (1982), 923–926.

[Str00] H. W. Strube, *The meaning of the Kelly-Lochbaum acoustic-tube model*, J. Acoust. Soc. Am. **108** (2000), 1850–1855.

[Sun87] J. Sundberg, *The Science of the Singing Voice*, Northern Illinois University Press, Dekalb, Illinois, 1987.

[Ter91] S. Ternström, *Physical and acoustic factors that interact with the singer to produce the choral sound*, Journal of Voice **5** (1991), no. 2, 128–143.

[Tit73] I. R. Titze, *The Human Vocal Cords: A Mathematical Model, Part I*, Phonetica **28** (1973), 129–170.

[Tit74] I. R. Titze, *The Human Vocal Cords: A Mathematical Model, Part II*, Phonetica **29** (1974), 1–21.

[Tit88] I. Titze, *A framework for the study of vocal registers*, Journal of Voice **3** (1988), 183–194.

[TMH+97] M. Tigges, P. Mergell, H. Herzel, T. Wittenberg and U. Eysholdt, *Observation and modelling glottal biphonation*, Acustica – acta acustica **83** (1997), 707–714.

[Väl95] V. Välimäki, *Discrete-time modeling of acoustic tubes using fractional delay filters*, Ph. D. thesis, Helsinki University of Technology, 1995.

[Val98] L. Valaczkai, *Atlas deutscher Sprachlaute*, Edition Praesens, Wien, 1998.

[vdB60] J. van den Berg, *Vocal ligaments versus registers*, Curr. Probl, Phoniat. Logoped, **1** (1960), 19.

[vdBZD57] J. van den Berg, J. Zantema and P. Doornenbal Jr., *On the air resistance and the Bernoulli effect of the human larynx*, J. Acoust. Soc. Am. **29** (1957), 626–631.

[Vil01] C. Vilain, *Modèle À Deux Masses Complet*, Personal communication, 2001.

[vK91] W. von Kempelen, *Mechanismus der menschlichen Sprache nebst der Beschreibung seiner sprechenden Maschine*, Wien, 1791.

[Vog95] H. Vogel, *Gerthsen Physik*, Springer Verlag, Berlin, Heidelberg, 1995.

[VPH+01] C. Vilain, X. Pelorson, A. Hirschberg, J. Willems and L. Le Marrec, *Study of the airflow through in-vitro oscillating pathological vocal folds*, Proceedings of the 2^{nd} International workshop on Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA) (Florence, Italy) (A. Manfredi, ed.), Uni Firenze, 2001.

[VPT99] C. Vilain, X. Pelorson and D. Thomas, *Effects of an induced asymmetry on the flow through the glottis in relation to voice pathology*, Proceedings 1^{st} Workshop on models and analysis of vocal emissions for biomedical applications (MAVEBA), 1999.

[WBH+90] B. Wein, R. Böckler, W. Huber, S. Klajman and K. Willmes, *Computersonographische Darstellung von Zungenformen bei der Bildung der Langen Vokale des Deutschen*, Ultraschall Med. **11** (1990), 100–103.

[WICT91] D. Wong, M. R. Ito, B. B. Cox and I. R. Titze, *Observation of perturbations in a lumped-element model of the vocal folds with application to some pathological cases*, J. Acoust. Soc. Am. **89** (1991), 383–394.

The most important functional components of the singing voice and their modeling are described. Exemplarily, overtone singing, different voice registers, and a voice with singer's nodules are synthesised. New methods for analysis of the singer's directivity and the vocal tract impedance at the mouth are presented.

Logos Verlag Berlin

ISBN 3-89722-997-8