ANALYSIS OF NASALS AND NASALIZED VOWELS BASED ON BRANCHED TUBE MODELS

K. Schnell, A. Lacroix

Institute of Applied Physics, Goethe-University Frankfurt
Robert-Mayer-Straße 2-4, D-60325 Frankfurt am Main
email: {Schnell, Lacroix}@iap.uni-frankfurt.de

ABSTRACT

The coupling of the nasal and mouth cavity at the velum is essential for the production of nasals and nasalized vowels. Therefore branched tube models are discussed and an algorithm is proposed to estimate the model parameters from speech signals. The tube model is realized by lattice structures in discrete time. The estimation algorithm is processed in two steps. The zeros are estimated starting with a general ARMA-estimation determining the parameters of the coupled tube branch. Hereafter the remaining model parameters are estimated by the minimization of the output power of the entire inverse filter, which is carried out iteratively. The analysis of nasal sounds and nasalized vowels shows that the algorithm is able to yield a good spectral fit between the magnitude response of the branched tube model and the speech spectrum.

1 INTRODUCTION

In the fields of speech analysis and production, branched tube systems are interesting particularly for nasals and nasalized vowels. The main cavities of the speech production system namely pharynx, mouth, and nose can be described by tube models. The modeling of the nasal tract itself by a single cavity is certainly a simplification of the actual situation, since the nasal tract is divided by the septum and has additional sinus cavities. However, in terms of the parameter estimation an unbranched nasal tract is favorable, since the system's ambiguity is reduced. The three single tube systems coupled at the velum are depicted in fig. 1. The coupled side branch represents the closed mouth

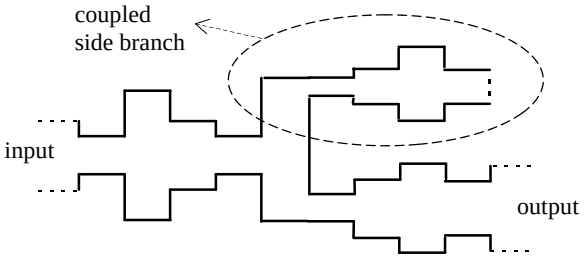

Figure 1: Branched tube system.

cavity in the case of nasals whereas for nasalized vowels the side branch describes the nasal tract. Hence it is possible to analyze nasals and nasalized vowels by this branched tube model. However, parameter estimation of branched tube models is complicated since the parameters of the side branch cause zeros in the transfer function and influence also the poles of the model.

Therefore the parameter estimation is done in two steps. At first the parameters of the side branch respectively the zeros are estimated, then the remaining parameters are determined by a specialized inverse filtering process. It has to be considered that the parameters of the side branch affect the poles of the transfer function. In [1] an identification procedure has been proposed for branched tubes where the poles of the model are estimated by the method of steepest descent minimizing an Euclidean cost function containing polynomial coefficients of the denominator of the transfer function in direct form. However, only a synthetic signal but no speech signals have been analyzed. In [2] an estimation procedure for branched tube models has been suggested which has been applied to speech signals, yet with the restriction that most of the parameters of the side branch must be known. In [3] an estimation procedure is described which allows the consideration of the output signals of the nostrils and lips simultaneously in the case of nasalized vowels, but this procedure needs a rather long computation time.

2 BRANCHED TUBE MODEL

The tube model, realized by lattice filters in discrete-time, describes the propagation of plane sound waves and consists of a concatenation of tube elements realized by delays which are interconnected by adaptors. Two-port adaptors describe the scattering of the sound waves at the tube junctions. A three-port adaptor with the scattering matrix $\boldsymbol{S}$

$$\boldsymbol{S}=\begin{pmatrix}\rho_2-1 & \rho_1 & 2-\rho_1-\rho_2\\ \rho_2 & \rho_1-1 & 2-\rho_1-\rho_2\\ \rho_2 & \rho_1 & 1-\rho_1-\rho_2\end{pmatrix},\qquad \rho_1=\frac{2A_1}{A_1+A_2+A_3}\quad \rho_2=\frac{2A_2}{A_1+A_2+A_3}.$$

represents the branching of a tube. The parameters ρ_1 and ρ_2 are functions of the three coupled tube areas A_i at the tube branch. To calculate the transfer function and to realize the inverse filter the 2×2 scattering transfer matrices $\boldsymbol{T}$ are used which describe the wave quantities at the left port as a function of the wave quantities at the right port. $\boldsymbol{T}_i$ is the scattering transfer matrix of a tube section with reflection coefficient r_i, combining the discontinuity of the cross section area and the uniform tube element

$$\boldsymbol{T}_i=\begin{pmatrix}1 & r_i z^{-1}\\ r_i & z^{-1}\end{pmatrix};\quad \begin{pmatrix}x_i^f\\ x_i^b\end{pmatrix}=\boldsymbol{T}_i\begin{pmatrix}x_{i+1}^f\\ x_{i+1}^b\end{pmatrix};\qquad (1)$$

x_i^f and x_i^b are the wave quantities of the forward and backward propagation in the tube. The three-port adaptor and the coupled side branch can be reduced to the 2×2 scattering transfer matrix

$\boldsymbol{T}_D$ which can be handled in the same manner as the $\boldsymbol{T}_i$ within the chain of matrices. The side branch is described by the transfer function $\tilde{H}(z) = Q/P$ as a ratio of two polynomials determined by the reflection coefficients of the side branch where the last reflection coefficient represents the termination. The reduced three-port adaptor shown in fig. 2 is given by

$$\boldsymbol{T}_D = \frac{1}{\rho_2(Q+P)} \cdot \begin{pmatrix} \langle \rho_1+\rho_2-1, 1\rangle & \langle \rho_2-1, 1-\rho_1\rangle z^{-1} \\ \langle 1-\rho_1, \rho_2-1\rangle & \langle 1, \rho_1+\rho_2-1\rangle z^{-1} \end{pmatrix}$$

with the abbreviation $\langle x, y\rangle := x \cdot Q + y \cdot P.$

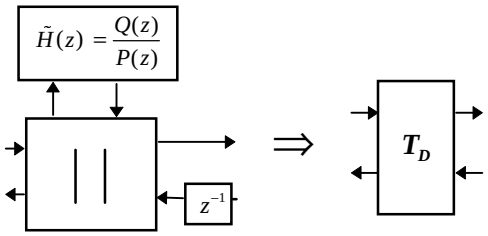

Figure 2: Reduced three-port adaptor with side branch.

$\boldsymbol{T}_D$ can be split into a matrix $\boldsymbol{T}'_D$ with polynomial elements and a common denominator polynomial B

$$\boldsymbol{T}_D = \frac{1}{B(z)} \cdot \boldsymbol{T}'_D, \qquad B(z) = \rho_2 B'(z) = \rho_2(Q+P).$$

The termination at the tube output is realized by a reflection coefficient R_e with a value about -0.9, describing approximately an open tube end with additional losses. If all scattering transfer matrices $\boldsymbol{T}_i$ and $\boldsymbol{T}_D$ are multiplied then the matrix $\boldsymbol{T}_c$ results and the transfer function of the entire tube system is given by:

$$H(z) = \frac{B(z)}{A(z)} = \frac{1}{T_c^{11} + R_e \cdot T_c^{12}} \quad . \qquad (2)$$

$H(z)$ depends on two elements of the matrix $\boldsymbol{T}_c$ and the termination R_e.The numerator B of $H(z)$ is the common denominator of the matrix $\boldsymbol{T}_D$.

3 PARAMETER ESTIMATION

The relationship between poles and zeros of the transfer function and the model parameters is not a one to one mapping, since the total number of poles and zeros is greater than the number of the tube parameters. This is caused by $\boldsymbol{T}_D$ respectively the side branch. So the tube parameters can be transformed into the poles and zeros, but not vice versa. However the zeros, represented by B, can be converted into the model parameters of the side branch. Therefore the estimation algorithm consists of two successive parts. One part for the zero estimation and the other part for the estimation of the poles of the tube system.

3.1 Estimation of the Zeros

The numerator of the transfer function is

$$B(z) = \sum_{i=0}^{L} b_i z^{-i} = \rho_2(1 + \sum_{i=1}^{L} b'_i z^{-i}) = \rho_2(Q(z) + P(z))$$

where the polynomials Q and P depend on the L reflection coefficients r_i of the side branch only. Thus it is possible to convert the coefficients b'_i into the parameters r_i of the side branch yielding Q and P. The conversion can be realized by a function, which is recursively defined. ρ_2 is disregarded in B, because it represents a gain factor only. The coefficients b'_i can be obtained by a general ARMA estimation. Additionally to the zeros also poles are estimated, because for an MA process the estimated zeros would model also the poles of the branched tube system. After the pole zero identification the estimated poles are ignored. For the ARMA estimation an iterative algorithm is chosen [4], which minimizes a spectral error derived from inverse filtering. The algorithm combines two known partial solutions of an AR and an MA analysis in an iterative way. The algorithm is explained in [4]. The order of the ARMA identification corresponds to the order of the transfer function of the tube system. After the pole zero estimation the polynomials Q and P can be obtained from the estimated coefficients b'_i. The estimated zeros affect the poles of the system which can be seen from $\boldsymbol{T}_D$. The remaining coefficients are estimated by inverse filtering.

3.2 Estimation of the Poles by Inverse Filtering

The concatenation of the scattering transfer matrices $\boldsymbol{T}_i$ and $\boldsymbol{T}_D$ can be interpreted as a signal flow graph which is the basis for the process of inverse filtering in the time domain. The inverse branched tube filter can be split into a purely recursive and nonrecursive part. The purely recursive part of the filter is the term $B'^{-1} = (Q+P)^{-1}$ in $\boldsymbol{T}_D$ which is known by the previous estimation of the zeros. The factor ρ_2 of B can be ignored in this case. So the purely recursive part of the inverse filter can be separated in a first step. This can be realized in the frequency domain with an inverse DFT yielding the signal x'

$$x' = \mathrm{IDFT}(X \cdot B'^{-1}) \qquad (3)$$

from the analyzed signal x. The nonrecursive part of the inverse filter is described by the sections $\boldsymbol{T}_i$ and the section $\boldsymbol{T}'_D$ which is depicted in fig. 3. For the inverse filter the matrices combine the quantities of the upper and lower path x_i^u and x_i^l by:

$$\boldsymbol{x}_i = \begin{pmatrix} x_i^u \\ x_i^l \end{pmatrix}; \quad \boldsymbol{x}_i = \boldsymbol{T}_i \cdot \boldsymbol{x}_{i-1}, \quad \boldsymbol{x}_{M+1} = \boldsymbol{T}'_D \cdot \boldsymbol{x}_M \quad .$$

The output of the nonrecursive part of the inverse filter is x_N^u and the input is given by x_0^u which is the signal x'. For the parameter

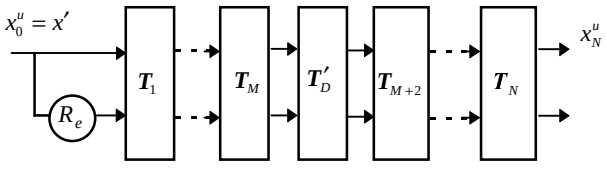

Figure 3: Flow graph of nonrecursive part of inverse filtering.

estimation the problem is to determine the parameters in a way, that the power of the output signal x_N^u is a minimum. In contrast to the Burg-method and related methods the output power should not be minimized after each section. Therefore each coefficient is estimated by minimizing the power of the output x_N^u, while the other coefficients are fixed. Only the sections behind the estimated

section $\boldsymbol{T}_i$ or $\boldsymbol{T}'_D$ affect the output power. So the two input signals x^u_{i-1} and x^l_{i-1} of the estimated section should be filtered with respect to the minimal power of the output x^u_N. The criterion for the minimum is:

$$\mathrm{E}\left[x_N^{u\,2}\right] \to \min. \quad \Rightarrow \quad \frac{\partial \mathrm{E}\left[x_N^{u\,2}\right]}{\partial r_i} = 0\,, \quad \frac{\partial \mathrm{E}\left[x_N^{u\,2}\right]}{\partial \rho_i} = 0 \qquad (4)$$

This approach is discussed in [5] for unbranched tube models. The parameters r_i and ρ_i have to be treated separately.

3.2.1 Estimation of Reflection Coefficients

r_i is the variable for the minimization (4), which results in the estimated reflection coefficient

$$\hat{r}_i = -\left(\overline{u_i^{11}\cdot u_i^{12}} + \overline{u_i^{11}\cdot l_i^{11}} + \overline{l_i^{12}\cdot u_i^{12}} + \overline{l_i^{12}\cdot l_i^{11}}\right) \Big/ \left(\overline{u_i^{12}\cdot u_i^{12}} + 2\cdot\overline{u_i^{12}\cdot l_i^{11}} + \overline{l_i^{11}\cdot l_i^{11}}\right) . \qquad (5)$$

Expected values $\mathrm{E}[x]$ are replaced by time averages $\overline{x}$. $u_i^{\lambda\beta}(n)$ and $l_i^{\lambda\beta}(n)$ are the filtered signals x^u_{i-1} and x^l_{i-1} behind $\boldsymbol{T}_i$

$$u_i^{\lambda\beta}(n) = f_i^{\lambda\beta}(n) * x^u_{i-1}(n)\,, \qquad l_i^{\lambda\beta}(n) = f_i^{\lambda\beta}(n) * x^l_{i-1}(n-1)\,.$$

The finite impulse responses $f_i^{\lambda\beta}(n)$ are determined by the coefficients of polynomials and can be derived from the matrix $\boldsymbol{F}_i$ which describes the tube sections between the output of the inverse filter and the estimated section

$$\boldsymbol{F}_i = \begin{pmatrix} F_i^{11} & F_i^{12} \\ F_i^{21} & F_i^{22} \end{pmatrix} = \begin{cases} \prod_{k=i+1}^{M} \boldsymbol{T}_k \cdot \boldsymbol{T}'_D \cdot \prod_{k=M+2}^{N} \boldsymbol{T}_k & \text{for } i = 1...M \\ \prod_{k=i+1}^{N} \boldsymbol{T}_k & \text{for } i = M+1...N-1 \end{cases} . \qquad (6)$$

3.2.2 Estimation of Parameters of the Three Port Adaptor

The two parameters ρ_i of section $\boldsymbol{T}'_D$ have to fulfill the criterion (4). If the output signal of the inverse filter is decomposed like

$$x_N^u = o + \rho_1 \cdot o_1 + \rho_2 \cdot o_2$$

then for the ρ_i we obtain the following formulas:

$$\hat{\rho}_1 = -\frac{\overline{o\cdot o_1} + \rho_2 \cdot \overline{o_2\cdot o_1}}{\overline{o_1\cdot o_1}}\,, \qquad \hat{\rho}_2 = -\frac{\overline{o\cdot o_2} + \rho_1 \cdot \overline{o_1\cdot o_2}}{\overline{o_2\cdot o_2}} \,. \qquad (7)$$

The signals o, o_1, o_2 can be calculated from the polynomials Q, P and polynomial matrix elements $F^{\lambda\beta}_{M+1}$ in (6) with

$$o_1 = u_Q^{11} - l_P^{11} + l_Q^{12} - u_Q^{12}\,, \qquad o_2 = l_Q^{12} + u_Q^{11} + l_Q^{11} + u_P^{12}\,,$$

$$o = -u_Q^{11} - l_Q^{11} - l_Q^{12} + l_P^{11} + u_P^{11} + l_P^{12} - u_P^{12} + u_Q^{12}\,,$$

and

$$u_Q^{\lambda\beta}(n) = q(n) * f_{M+1}^{\lambda\beta}(n) * x_M^u(n)\,,$$

$$l_Q^{\lambda\beta}(n) = q(n) * f_{M+1}^{\lambda\beta}(n) * x_M^l(n-1)\,,$$

$$u_P^{\lambda\beta}(n) = p(n) * f_{M+1}^{\lambda\beta}(n) * x_M^u(n)\,,$$

$$l_P^{\lambda\beta}(n) = p(n) * f_{M+1}^{\lambda\beta}(n) * x_M^l(n-1)\,.$$

$q(n)$ and $p(n)$ are determined by the polynomial coefficients

$$Q = \sum_k q(k)\cdot z^{-k} \quad \text{and} \quad P = \sum_k p(k)\cdot z^{-k}\,.$$

3.2.3 Iterative Procedure

$\hat{r}_i$ and $\hat{\rho}_i$ by (5) and (7) are the optimal parameters on condition that the other coefficients are given. Since no priori information is available, the algorithm starts with all parameters r_i equal to zero and ρ_i equal to $2/3$ representing a "neutral position". Then the parameters are estimated one after the other by equations (5) and (7) from r_1 till r_M, ρ_1, ρ_2 and r_{M+2} till r_N completing one iteration of the algorithm. Since the solution of one parameter depends on the other parameters, the next iteration improves the estimate because the $\hat{r}_i$ and $\hat{\rho}_i$ of the last iteration are used for the calculation of $\boldsymbol{F}_i$ in the next iteration. The values of the estimated parameters are restricted by physical constraints, to guarantee positive cross section areas. The tube termination as well as the length of the tubes have to be set before the algorithm starts.

3.3 Analysis of Periodic Signals

A single period is analyzed on the assumption that the signal values outside of the analysis segment are determined by periodic continuation. On this assumption a time delay transforms a period, represented as a vector $\boldsymbol{x}$ of length N

$$(x_1, x_2, ..., x_{N-1}, x_N)^{\mathsf{T}} \quad \text{into} \quad (x_N, x_1, x_2, ..., x_{N-1})^{\mathsf{T}}\,,$$

so that no signal values get lost by processing. By this definition of the time delay the estimation algorithm is independent of the phase of the analyzed signal to a great extent.

4 ANALYSIS OF SPEECH SOUNDS

Signals from prescribed branched tube systems have been analyzed to test whether the algorithm is able to reach the optimal solution. The analysis of these test signals shows, that the algorithm converges close to the optimal solution after several iterations.

For speech signals the influence of the excitation and radiation of voiced speech is separated from the vocal and nasal tract by an adaptive preemphasis. This is carried out by a repeated Burg method of first order. For the analysis of voiced speech signals it should be considered, that adjacent periods differ slightly and also the results of the analysis. To avoid this effect, few periods are averaged in the spectral domain to produce a single segment.

4.1 Analysis of Nasals

The sampling rate of the analyzed nasal /n/ is 16 kHz. Figure 4 shows the estimated magnitude responses in comparison with the

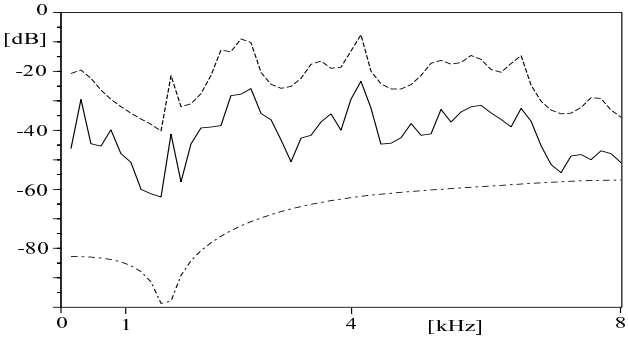

Figure 4: Analysis of /n/ with preemphasis: Estimated magnitude response after 50 iterations (top solid line), DFT of analyzed speech signal (dashed-dotted line), estimated zeros of the model (bottom dashed line).

spectrum of the analyzed nasal /n/ prefiltered by an adaptive preemphasis. Furthermore in fig. 4 the contribution of the zeros is plotted. The number of tubes are chosen corresponding to the actual tract lengths. Hence the nasal tract consists of 13 tubes and the pharynx has 8 tubes. For the mouth cavity 5 tubes are provided for the nasal /n/. The termination at the nostrils has the value -0.9.

4.2 Analysis of Nasalized Vowels

The nasalized vowels / $\tilde{a}$ / and / $\tilde{i}$ / with a sampling rate of 16 kHz are analyzed. In the case of nasalized vowels [6] the side branch represents the influence of the nasal tract in comparison to the analysis of nasals. The termination at the lips has the value -0.9. The estimated magnitude responses are shown in figures 5 and 6. The estimated vocal tract areas of the nasalized / $\tilde{a}$ / are depicted

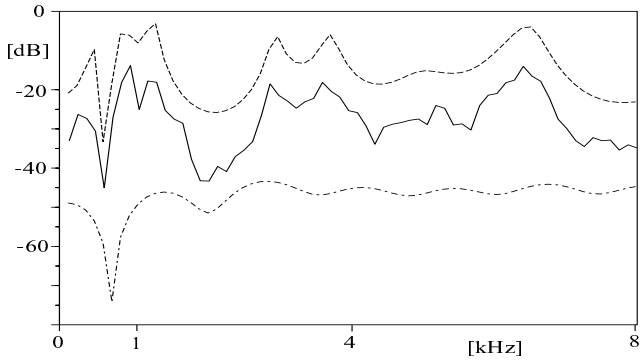

Figure 5: Analysis of nasalized / $\tilde{a}$ / with preemphasis: Estimated magnitude response after 50 iterations (top solid line), DFT of analyzed speech signal (dashed-dotted line), estimated zeros of the model (bottom dashed line).

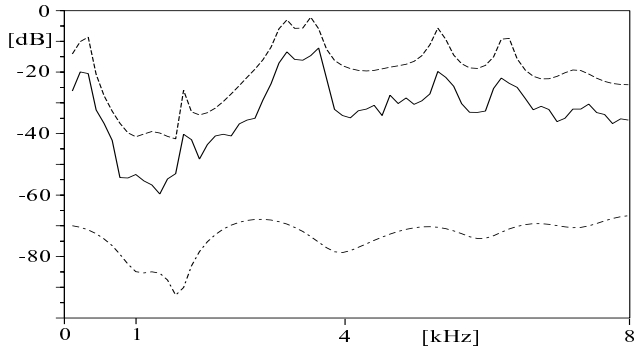

Figure 6: Analysis of nasalized / $\tilde{i}$ / with preemphasis: Estimated magnitude response after 50 iterations (top solid line), DFT of analyzed speech signal (dashed-dotted line), estimated zeros of the model (bottom dashed line).

in fig. 7 showing a reasonable shape. The estimated areas of the side branch respectively the nasal tract are shown in fig. 8 for nasalized / $\tilde{a}$ / and / $\tilde{i}$ /. The estimated areas of the nasal tract are similar for the two different vowels. The intensity of the coupling of the nasal tract at the velum affects the locations and strength of the zeros, since the first areas of the nasal tract at the velum depend on the velum position. The side branch is coupled to the vocal tract at tube section number 9. Although the determined parameters of the side branch restrict the pole configuration, the poles are estimated very well by the iterative inverse filtering procedure which can be seen in figs. 4-6. The estimated magnitude responses are in very good coincidence with the DFT-spectra.

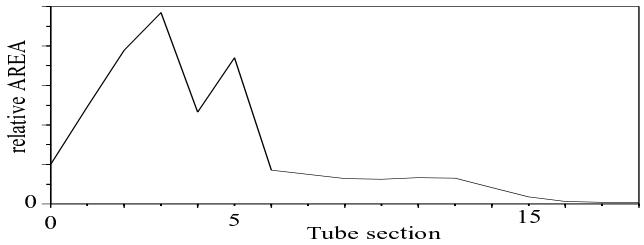

Figure 7: Estimated vocal tract areas from the nasalized vowel / $\tilde{a}$ /. At the left side the lips are located.

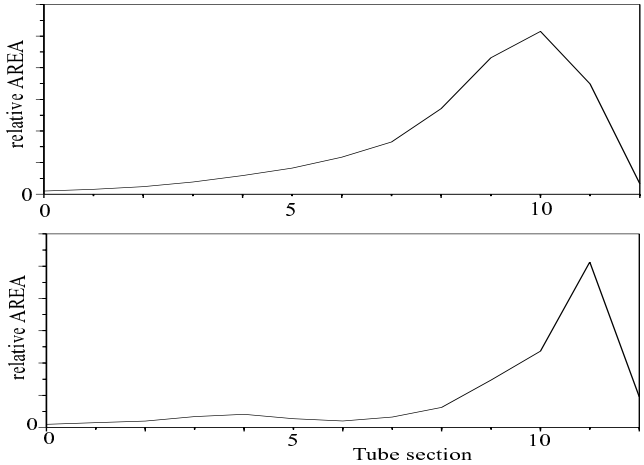

Figure 8: Estimated areas of the side branch respectively the nasal tract. At the left side the velum is located. For the nasalized vowels / $\tilde{a}$ / (top) and / $\tilde{i}$ / (bottom).

5 CONCLUSION

The proposed estimation algorithm has been applied to nasals and nasalized vowels. Examples of analyzed speech signals show that the speech spectra are modeled by the magnitude responses of the estimated branched tube systems very well. The shape of the estimated vocal tract areas is reasonable.

6 REFERENCES

[1] Lim I.T.; Lee B.G. :“Lossy Pole-Zero Modeling of Speech Signals“, *IEEE Trans. Speech and Audio Processing*, Vol. 4, No. 2, pp. 81-88, March 1996.

[2] Liu, M.; Lacroix, A.: “Improved Vocal Tract Model for the Analysis of Nasal Speech Sounds“, *Proc. ICASSP-96*, Atlanta, USA, pp. 801-804, 1996.

[3] Schnell, K.; Lacroix, A*:* „Parameter Estimation from Speech Signals for Tube Models“, *Proc. Joint EAA/ASA meeting, Forum Acusticum 1999*, Berlin, Germany, on CD-ROM, 1999.

[4] Schnell, K.; Lacroix, A.: “Pole Zero Estimation from Speech Signals by an Iterative Procedure“, *Proc. ICASSP-2001*, Salt Lake City, USA, 2001.

[5] Schnell, K.; Lacroix, A.: “Inverse Filtering of Tube Models with Frequency Dependent Tube Terminations“, *Proc. EUROSPEECH-2001*, Aalborg, Denmark, pp. 2467-2470, 2001.

[6] Feng, G.; Castelli, E.:“Some acoustic features of nasal and nasalized vowels: A target for vowel nasalization“, J.Acoust.Soc.Am., Vol. 99, No. 6, pp. 3694-3706, June 1996.