ACOUSTIC MODEL FOR AN ARTICULATORY-FORMANT SPEECH SYNTHESIZER

A. Miller [1] *and V. Sorokin*
Institute for Information and Transmission Problems, Moscow

ABSTRACT

Effective computational methods for solving the vocal tract equation relative to resonance frequencies, amplitudes and dampings were developed. Effects of the yielding walls were described with the equivalent sound velocity which enabled an accurate frequencies-domain parameters calculation. A nasal cavity coupling was also computed. Validity of the acoustic model was tested in synthesis experiment.

INTRODUCTION

In an articulatory-formant speech synthesizer acoustic processes are presented by a set of frequency-domain parameters - formants, dampings and amplitudes coefficients. There was tried on two approaches for these parameters search. In [1] formant parameters were calculated by converting transfer function derived through the transmission line analog representation of the vocal tract. In method of [2] formants was obtained directly from the vocal tract equation. Potentially the last approach takes advantage of more accurate and numerically economical description of the vocal tract acoustics because of applying direct analytical consideration of the vocal tract equation.

Following second approach we attempted to develop algorithms for formant parameters calculation from the vocal tract area- function. Algorithms are based on frequency-domain analysis of the vocal tract equation and designed to include walls vibrating, distributed losses and nasality. The paper is organized as follows. First, algorithms for vocal tract computation are presented. Then effect of nasal coupling will be described. Then practical realization of acoustic model will be discussed. Finally, obtained results are summarized.

[1]Currently with SR Telecom Inc., Montreal, Canada

SEARCH OF FORMANTS

As usual we assume that wave propagation in lossy acoustic tube with soft walls is described as it was done in [3][4]. For formants search, first of all, the appropriate expressions to account for wall yielding are to be found. To account for yielding walls we used the method of an effective sound velocity [5]. For small perturbations of area function $A_0(x,t)$ after wall displacement in the radial direction we get equivalent sound velocity as

$$c_e(t) = \left(\rho_0\beta_0 + \frac{S_0\xi}{pA_0}\right)^{-0.5}$$

where $\beta_0 = 1/\rho_0 c_0^2$.

In case of quasi steady-state deformations of the vocal tract frequency decomposition of Webester's equation gives us the following equation for spatial mode

$$(A_0\varphi')' + \frac{\omega_k^2 A_0}{c_e^2}\varphi = 0 \qquad (1)$$

where $\varphi(x)$ is the spatial component of the pressure, respectively, " $'$ " and " $''$ " denote first derivative on x. For a cylindrical tube with vibrating walls

the equivalent sound velocity c_e becomes

$$c_e^2(x) = \frac{c_0^2}{1 - 2Y_w/\mathrm{j}\omega_k a\beta_0} \qquad (2)$$

where $a(x)$ is radius of the tube, $Y_w(\omega_k, x)$ is conductivity of walls.

The equivalent sound velocity appears to be a generalization of effective sound velocity of [6] for the case of arbitrary dependence of the wall impedance from vocal tract geometry. Besides, Eq. (1) takes into account the active resistance ignored in [6].

The boundary condition at the lips end following model of [3] is $\varphi'(L) + 3\pi\varphi(L)/8a(L) = 0$. The boundary condition at the glottis for the closed vocal slit is $\varphi'(0) = 0$.

The eigenfrequencies of Eq. (1) can be found by solving an equivalent initial value problem (Kauchy problem) by means of shooting method. However, the original boundary problem represented as the equivalent Kauchy problem has unpleasant error propagation properties when abruptness occurs in the area function $A_0(x)$. To provide robustness the method of a phase function [7] was implied. This method does not superimpose explicit constraints on the spatial derivative of $A_0(x)$ and, besides, it offers more economical computation.

Table 1 represents three first formant frequencies calculated for the area functions of Russian vowels taken from [8](column E). Column A represents absolutely rigid walls. The cross-section shape was described as ellipse with the axis ratio 2, the wall impedance was $Z_w = 800 + \mathrm{j}1.3\omega$ g/cm·s . The formant frequencies represented in the column B were taken from [6], calculated by the effective sound velocity method. As one can see, essential errors can be observed for the vowels /a/, /i/ and /i/ in relation to the measured frequencies. Meanwhile, the frequencies calculated with Eq. (1) (the column D) are close to those calculated by the transmission line analog (the column C) and both are better fit to the experimental data (the column E).

F1	A	B	C	D	E
a	636	668	677	679	700
e	424	459	459	442	440
o	495	538	538	536	535
u	228	302	291	286	300
i	231	295	285	257	240
ɨ	289	345	-	312	300
F2					
a	1098	1101	1088	1105	1080
e	1980	1981	1998	2003	1800
o	848	887	869	875	780
u	606	629	631	605	625
i	2245	2284	2338	2329	2250
ɨ	1531	1541	-	1562	1480
F3					
a	2400	2477	2412	2488	2400
e	2741	2821	2748	2887	2550
o	2380	2398	2432	2386	2500
u	2385	2391	2450	2385	2500
i	2888	3102	2898	3116	3200
ɨ	2414	2421	-	2366	2230

Table 1: Resonance frequences for Russian vowels.

DAMPINGS

Acoustic losses may be presented as a resistance $R(\omega, x)$ reflecting losses arising from viscous friction and lips radiation and an admittance $Y(\omega, x)$ reflecting losses caused by the thermal conduction and wall vibrations [5]. Those losses are accounted for as an effective coefficients of damping δ_k for each k-th mode, so that the equation for temporal mode is

$$\psi'' + 2\delta_k\psi' + \omega_k^2\psi = 0.$$

Using the decision of the last equation and Eq. (1) we get after integrating

by x and averaging by $T = \omega_k/2\pi$ as follows

$$\delta_k = \int Y\varphi_k^2 dx + \int R(A_0\varphi_k/c_e)^2 dx$$

Expression for δ_k was obtained under assumption of small losses, i.e. $\delta_k/\omega_k \ll 1$.

AMPLITUDES

The amplitude of forced oscillations is determined as

$$\psi'' + 2\delta_k\psi' + \omega_k^2\psi = e_k,$$

where the coefficient e_k is calculated as a result of decomposition of the excitation function $G(x,t)$ with the eigenfunctions of Eq. (1) as follows

$$e_k = \int G(x,t)A_0\varphi_k/c_e^2(x)dx.$$

For the source of current in the vocal slit function of excitation becomes $G(0,t) = u_g'(0,t)$. So the amplitude of oscillations determined as

$$e_k(t) = u_g'(t)A_0(0)\varphi_k(0)/c_e^2(0)$$

Other sources of excitation are determined in the same way.

NASALIZATION

If the velum is lowered then the vocal tract consists of three united parts: pharyngeal, oral and nasal cavities. If eigenfunctions for all the cavities are known, then resonance frequencies can be obtained from the condition of inconsistency of flow and pressure at the velum port [9].

Amplitudes and dampings are calculated in the manner analogous to non-branched case described above.

The resonance frequencies for the vocal tract with the coupled nasal branch are represented in Table 2 for the area functions of vowels used in Table 1. The velum opening was equal to 1 cm. To search formant frequencies we used numerical algorithm developed for branched vocal tract in [10] . As it was expected, some additional resonances appear between the resonances inherent to the vocal tract itself.

	F1	F2	F3	F4	F5
a	573	912	1157	2341	2961
e	449	906	1956	2638	3020
o	495	888	2319	2775	3612
u	314	537	758	2273	2630
i	315	881	2325	2737	3146
y	347	876	1557	2353	2744

Table 2: Resonance frequences for nasalized Russian vowels.

PARALLEL SYSTEM

The vocal tract model provides computation of amplitudes and bandwidths which are presented as a vector

$$\vec{x} = \left[\vec{F},\ \vec{B},\ \vec{\varphi}_L,\ \vec{\gamma}_N,\ \vec{A}\right]^{\mathrm{T}},$$

where $\vec{F} = [F_1, ..., F_m]^{\mathrm{T}}$, m is number of formants, the rest of the $\vec{x}$ components are defined in the same way. Variations of articulatory and acoustic conditions may cause variations of quantity of the formants for given frequency band. Selectors were introduced into parallel system to provide automatic ordering of calculated formants. The selector and the block of formant filters constitute two parallel branches for independent processing of nasal and oral resonances.

The selector compares two successive sets of calculated formants and provide distribution of calculated formant parameters within fixed set of formant filters. We used eight canals for each branch to cover the 5 kHz band.

The model of the synthesizer embodies an aerodynamic model for computation function of excitation of the vocal tract. Consideration of aerodynamic model is beyond of the paper and may be found elsewhere [1].

Now, let us estimate the number of operations required for frequency parameters computation. For 5 ms interval of area function update and 0.5 cm of spatial increment (error of formant estimation less the 4%) it takes about 1.5sec of IBM PC/486/66 CPU time per 1sec of synthesized speech. Computational expenses for the formant frequency calculations are about 15% of the computational cost for the speech wave computation with sampling frequency 20 kHz.

To verify validity of the described acoustic model, we had a synthetic experiment with an articulatory-formant speech synthesizer based on the aerodynamic model of the vocal tract [9] and laboratory model of the vocal tract dynamics. Listening tests with speech material consisting of words and short phrases indicated that both intelligibility and naturalness of synthesized expressions were close to those of natural ones.

CONCLUSIONS

Combining the advantages of both analytical and numerical approaches, the effective technique for vocal tract computation has been developed. Simple and fast algorithms for calculation of the main acoustic parameters - frequencies, amplitudes and dampings of resonance oscillations - take into account such important factors as yielding walls, nasalization and distributed losses.

Algorithm for formants search provides the amount of operations about 1.5 s of IBM PC/486/66 CPU time per sec with no special signal processing equipment. The achieved accuracy and speed are acceptable for application of these computational algorithms in the tasks of articulatory speech synthesis.

References

[1] Lin Q. "Theory of speech production and articulatory speech synthesis," Ph.D. Thesis, KTH, Stockholm, 1990.

[2] Coker C.H. "A model of articulator dynamics and control," Proc. IEEE 64, 1976, 452-460

[3] Morse P.M. Vibration and sound (McGraw-Hill, New York), 1948

[4] Portnoff M. "A quasi one dimensional digital simulation for the time-varying of the vocal tract," MS Thesis, MIT, 1973.

[5] Flanagan J.L. "Speech Analysis, Synthesis and Perception (Springer Verlag Berlin, Heidelberg, New-York), 1972.

[6] Badin P., Fant G. "Notes on vocal tract computation," STL/QPSR 6, 1984, 53-108.

[7] Hall G. and Watt J.M. Modern numerical methods for ordinary differential equations (Clarendon Press, Oxford), 1976.

[8] Fant G. Acoustic Theory of Speech Production (Mouton, The Hague, The Netherlands), 1960.

[9] Sorokin V.N. Theory of speech production. Moscow, 1985. (in Russian)

[10] Miller A.H., Sorokin V.N. "Methods of shooting for the vocal tarct equation", Acoustical journal, 2, 1991, 361-367. (in Russian)