ELECTRONICS AND ELECTRICAL ENGINEERING
ISSN 1392 – 1215 2009. No. 8(96)
ELEKTRONIKA IR ELEKTROTECHNIKA

SIGNAL TECHNOLOGY
T 121
SIGNALŲ TECHNOLOGIJA

Simulation of Spectral Movement of Vowels using Acoustical Tube

D. Balbonas, G. Daunys
Department of Electronics, Siauliai University,
Vilniaus str. 141, LT-76353 Šiauliai, Lithuania; phone: +370 41 595835; e-mail: dainius@tf.su.lt

Introduction

The study of speech spectrum and vocal tract is important to voice technologies used in human-computer interaction. One of these technologies is speech synthesis. Speech synthesis is concerned with many disciplines such as: signal technologies, communication and information theory, linguistics, physiology and computer science [1].

The research is related with problems of speech synthesis and possibilities to use electronic circuits for solving these problems. The most actual areas connected with speech synthesis are: aid for disables, medical applications, learning activities, telecommunications and computer assisted (administration and service) work places.

Speech synthesis allows for disabled (visually impaired, with motional disorders) to access computer-aided job places and various information systems. Speech synthesis in learning activities can help to improve learning process or can help in speech therapy. The largest market where speech synthesis can be useful is communication and administration facilities [2].

Earlier author's articles also were related to spectral movement of vowels [3, 4]. In articles [3, 4] authors found out the most significant parameters which influence spectral movement of vowels.

Model of speech tract is described in [5, 6]. One of it's practical realizations in Mathworks MATLAB is given by Childers [7].

Time domain methods of speech signal production allow to take into account more details and simulated speech signal is more precise. In our research the cascade connection of many T form LCR circuit was used to construct the speech tract filter. Integration of a system of differential equations was used to obtain speech signal, which was used for formants analysis. The results were compared with formants obtained using two-ports model of the vocal tract.

Method

The model of electronic acoustical tube consists of two parts. The first part is excitation model and the second speech tract model. Excitation model is shown on Fig. 1, speech tract model is shown on Fig. 2.

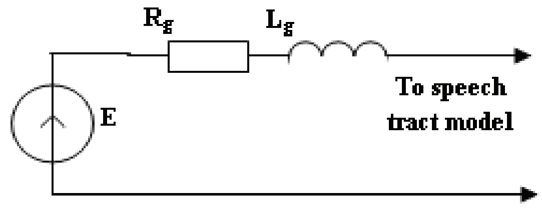

Fig. 1. Excitation model

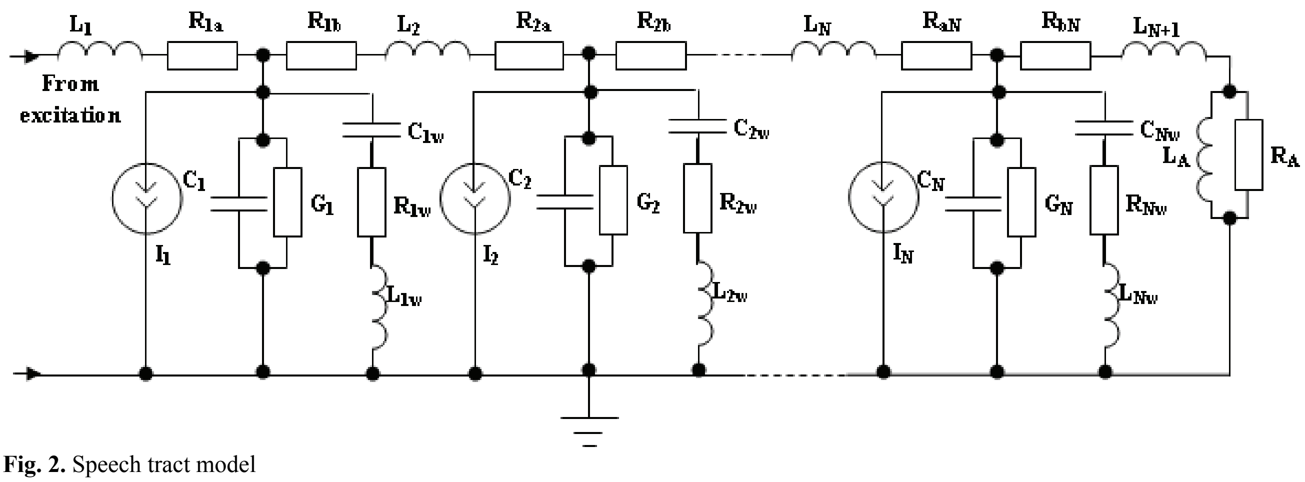

Fig. 2. Speech tract model

Voltage source in excitation was used for modeling atmospheric pressure. Current sources in T segments used to incorporate extra air flow because of movement of vocal tract pieces.

RLC circuit arranged parallel to main capacity used incorporate movement of vocal tract walls

Relation between speed of air volume (analogue electric currency) and pressure of air (analogue of voltage) can be interpreted via impedance Z using resistance, capacity, and inductance. The mass of air is compressible and inertial. Compressibility can be equated to capacity and inertness can be equated to inductance. The capacity, inductance of A area tube and load resistance are described in paper [8].

The model shown on Fig. 2 and described in this paper can be used find out transition of areas of cross-section of speech tract during transition between vowels.

The mathematical model of acoustical tube consists of state variable method, where variable are voltage in capacity (correspond pressure of sound) and current in inductance (corresponds rate of air volume). Every one segment has the first order differential equation. Coefficients on the right side are dependable on time. The changing of coefficient is slow because of slow excitation.

Using Kirchoff's rules the equations of currency for nodes and equations of voltage for loops were wrote.

Voltage on inductance can be denominated as derivative of inductance current multiplied by inductance also current in capacity can be dominated as derivative of voltage on capacity multiplied by capacity. Using these denominatives all equations were rewritten, see equations (1) and (2)

$$\begin{cases} L_1 \dfrac{di_{L_1}}{dt} = E - u_{R_g} - u_{R_{1a}} - u_{C_1}, \\ L_2 \dfrac{di_{L_2}}{dt} = u_{C_1} - u_{R_{1b}} - u_{R_{2a}} - u_{C_2}, \\ L_{1w} \dfrac{di_{L_{1w}}}{dt} = u_{C_1} - u_{C_{1w}} - u_{R_{1w}}, \\ L_3 \dfrac{di_{L_3}}{dt} = u_{C_2} - u_{R_{2b}} - u_{R_{3a}} - u_{C_3}, \\ L_{2w} \dfrac{di_{L_{2w}}}{dt} = u_{C_2} - u_{C_{2w}} - u_{R_{2w}}, \\ \dots\dots\dots\dots\dots\dots \\ L_N \dfrac{di_{L_N}}{dt} = u_{C_N} - u_{R_{bN}} - U_A, \\ L_{Nw} \dfrac{di_{L_{Nw}}}{dt} = u_{C_N} - u_{C_{Nw}} - u_{R_{Nw}}. \end{cases} \tag{1}$$

Equations (1) and (2) can be directly integrated, but such integration gives unstable solutions. On this situation indirect integration was choused; derivatives were changed into finite alteration also was used averaging of state variables of neighboring time moments. The future values of time moment calculating using system of linear equations. The system of linear equations is simplified and reorganized

The matrix of system of linear equations and column of free members created from equation system.

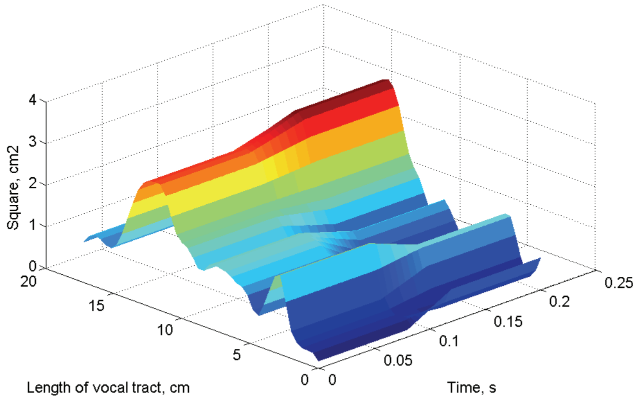

Fig. 3. Transition of of cross-section areas between short and long Lithuanian vowel [i]

$$\begin{cases} C_1 \dfrac{du_{C_1}}{dt} = i_{L_1} - i_{G_1} - i_{L_2} - i_{L_{1w}} - i_1, \\ C_{1w} \dfrac{du_{C_{1w}}}{dt} = i_{L_{1w}}, \\ C_2 \dfrac{du_{C_2}}{dt} = i_{L_2} - i_{G_2} - i_{L_3} - i_{L_{2w}} - i_2, \\ C_{2w} \dfrac{du_{C_{2w}}}{dt} = i_{L_{2w}}, \\ \dots\dots\dots\dots\dots\dots \\ C_N \dfrac{du_{C_N}}{dt} = i_{L_N} - i_{G_N} - i_{L_{N+1}} - i_N, \\ C_{Nw} \dfrac{du_{C_{Nw}}}{dt} = i_{L_{Nw}}, \\ \dfrac{U_A}{R + R_{bN}} = i_{L_{N+1}}. \end{cases} \tag{2}$$

The most important solution of linear equations is next-to-last. This solution indicates the air flow in the output of speech tract and the first order derivative according time is speech signal in time domain.

Also this model gives opportunity to simulate transition of cross-section squares between two vowels. In Fig. 3 is shown transition of cross-section areas of vocal tract between short and long Lithuanian vowel [i].

Also a cascaded two-port was used to describe the acoustical tract.

The algorithm using two-port **A** parameters was proposed. **A** parameters were calculated using acoustic tube section electrical T form model (Fig. 4).

In equation 3 is shown calculation of C and L

$$L = \frac{\rho}{S} \cdot \Delta x, \qquad C = \frac{S}{\rho c^2} \Delta x, \tag{3}$$

here, ρ – density of air, S- cross-section of tube section, Δx – length of tube section, c – velocity of sound in air.

The A parameters for network shown in equation 4:

$$a_{11} = 1 + \frac{(R_1 + 0{,}5\,jwL_1)(Z_n + Z_l)}{Z_n Z_l},$$

$$a12 = R_1 + jwL_1 + R_2 +$$
$$+ \frac{(R_1 + 0{,}5\,jwL_1)(R_2 + 0{,}5\,jwL_1)(Z_n + Z_l)}{Z_n Z_l}),$$

$$a21 = \frac{Z_n + Z_l}{Z_n Z_l},$$

$$a22 = 1 + \frac{(R_2 + 0{,}5\,jwL_1)(Z_n + Z_l)}{Z_n Z_l}. \quad (4)$$

here

$$Z_l = \frac{1}{jwC_1G_1\left(\frac{1}{jwC_1} + \frac{1}{G_1}\right)}, \; Z_n = R_2 + jwL_2 + \frac{1}{jwC_2}.$$

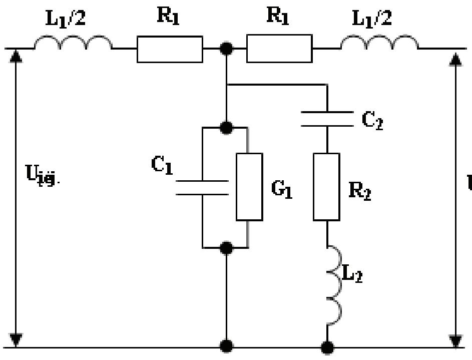

Fig. 4. T form model of acoustical tube section

Impedance of load makes influence on frequency characteristic of acoustical tube. Impedance of load consists of parallel connected inductance L_r and resistance R. Values of L_r and R found from equation below[7]:

$$R = \frac{128\rho c}{9\pi^2 S}, \quad L_r = \frac{8\rho}{3\pi\sqrt{\pi S}}. \quad (5)$$

Amplitude frequency response was calculated using sing A parameters of cascaded two-ports

Results and discussion.

Using an electronic acoustical tube model the transition from the short to long vowels in time domain and frequency domain were simulated.

In table 1 is shown the frequencies of first five formants and their standard deviations.

In Fig. 5 simulation results of “a” vowel, when simulating the transition from the short to long vowel in time domain is shown. In Fig. 6 and Fig. 7 simulation results of “a”, and “u” vowels, when simulating the transition from the short to long vowel in frequency domain are shown. Formant variation in a continuous line derived from the use of modeling two ports. Roller marked formant values obtained from simulated time domain signals using LP analysis implemented in Praat program.

Table 1. Frequencies of simulated formants and standard deviations of stationary vowels

Vowel	F_1, Hz	F_2,Hz	F_3, Hz	F_4, Hz
a	509.2±8.1	1293.6±3.3	2437.7±3.6	3446.4±17.6
a:	668.4±1.3	1081.3±7.6	2378.9±8.5	3398.9±26.1
e	487.6±2.3	1396.7±11.8	2293.3±5.6	3503.1±1.5
e:	595.9±8.6	1471.3±5.0	2281.8±9.3	3358.5±16.0
i	398.9±5.7	1551.6±9.9	2297.6±3.5	3500.5±8.2
i:	405.1±5.9	1588.0±1.5	2354.4±3.1	3617.7±6.9
o:	578.4±2.1	934.8±1.6	2365.6±18.9	3315.8±28.5
o	401.7±5.4	1000.1±1.0	2414.2±24.8	3399.9±7.9
u:	378.3±2.0	1128.5±8.2	2221.6±19.8	3609.3±45.1
u	294.9±5.3	891.0±3.8	2151.4±35.8	3522.6±145.8

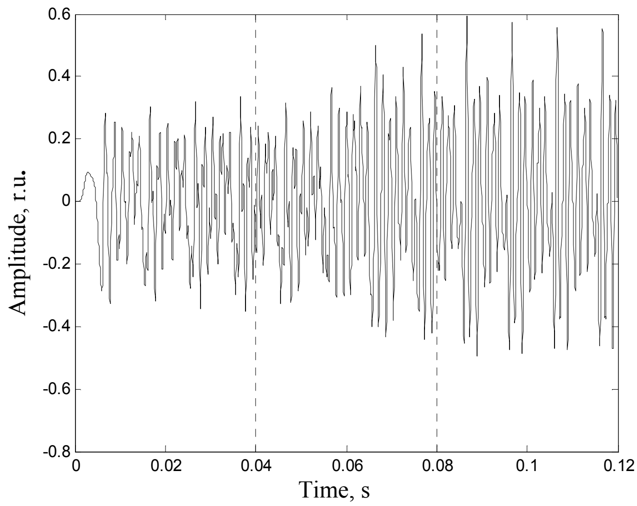

Fig. 5. Transition between short and long vowel „a“

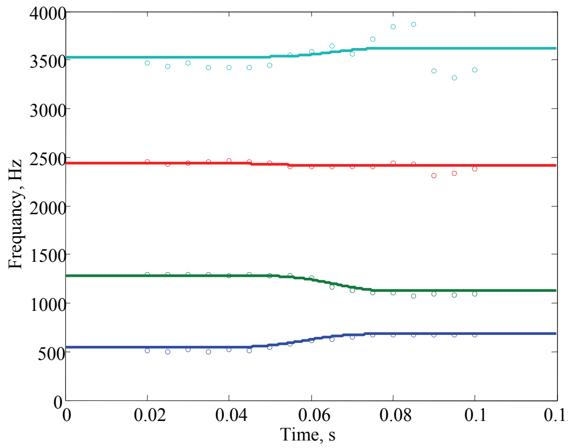

Fig. 6. Transition between short and long vowel „a“ (Formant from bottom F_1, F_2, F_3, F_4)

The beginning of transition is at 0.04s and the end of transition is at 0.08s. During transition from a short to long vowel in time domain the change of amplitude is observed, and in frequency domain change of formant frequency is observed.

The same result of simulation of transition for both methods achieved in vowel “a”. For other vowels some difference results between methods were observed. In Fig. 5 the error of formant determination using LP method is observed. The frequency of first formant was too low and was ignored, so the formants were shifted up by one number.

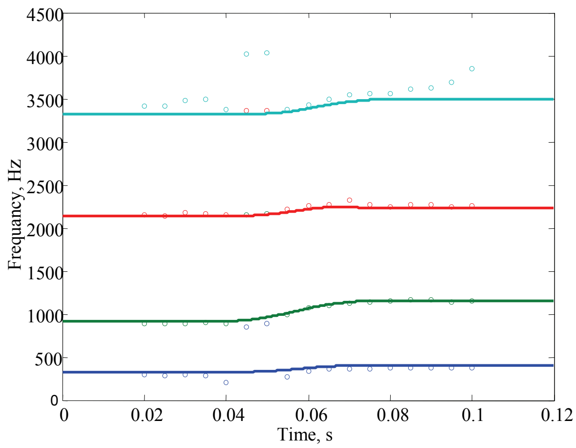

Fig. 7. Transition between short and long vowel „u" (Formant from bottom F_1, F_2, F_3, F_4)

Conclusions

Two methods were used to simulate formants movement in vowels. Speech signal simulation is more accurate in time domain. However there is no easy way to extract formants values from the signal. A LPC method was used to obtain formants frequencies. A method, which describes an acoustic tube by serially connected two-ports, indicates better results for formant movement evaluation because a frequency response can be calculated directly.

References

1. **Rabiner L., Juang B.–H.** Fundamentals of speech recognition. – Prentice–Hall International. – 1993. - ISBN 0–13–285826–6. – P. 277.
2. **Rudžionis A., Ratkevičius K., Rudžionis V., Kasparaitis P., Šalna B.** Projekto "Lietuvių kalbos bazinių elementų diegimas kompiuterinėse sistemose" skyriaus „Balso technologijų taikymo lietuvių kalbai analizė ir perspektyvinių veiklos krypčių pagrindimas" atasakita // Prieiga per internetą http://www.likit.lt/frames/balso_tech/balsotech_st.htm [2008 05 23].
3. **Balbonas D., Daunys G.** Movement of Formants of Vowels in Lithuanian Language // Electronics and Electrical Engineering. – Kaunas: Technologija, 2007. – No. 7(79). – P. 15–18.
4. **Balbonas D., Daunys G.** Movement of Formants of Sound *a* in Lithuanian Language // Electronics and Electrical Engineering. – Kaunas: Technologija, 2006. – No. 8(72). – P. 75–78.
5. **Shaughnessy O. Speech D.** Communications. Human and Machine. – New York: IEEE PRESS. 2000. 547 P. ISBN 0–7803–1111–6.
6. **Deller J. R., Hansen J. H. L., Proakis J. G.** Discrete–Time Processing of Speech Signals. – New York: IEEE PRESS. 2000. 908 P. ISBN 0–7803–5386–2.
7. **Childers D. G.** Speech processing and synthesis toolboxes. John Wiley & Sons, 2000, 484 p.
8. **Balbonas D., Daunys G.** Lietuvių kalbos balsių spektro modeliavimas akustiniu vamzdžiu // Biomedicininė inžinerija 2008, Kaunas: Technologija. P. 249–252. ISBN 978–9955–25–576–5

Received 2009 06 22

D. Balbonas, G. Daunys. Simulation of Spectral Movement of Vowels using Acoustical Tube // Electronics and Electrical Engineering. – Kaunas: Technologija, 2009. – No. 8(96). – P. 85–88.

A vocal tract can be described using the serially connected electrical circuits of T-type. Its frequency response was obtained by two ways: simulation of the signal and later calculation of the spectrum, and calculation of frequency response of serially connected two-ports. In the first case, a mathematical model of the acoustic tube was built using a state variable approach, where the state variables are capacitor voltages (equivalent to sound pressure) and the inductance currents (equivalent to air volume speed). An obtained system of linear differential equations was integrated by indirect method. In the second case, two-ports "a" parameters were used to obtain all system frequency response. During vowel pronunciation the areas of the vocal tract cross-section are obtaining changes over time. In both methods formant trajectories were extracted from the vocal tract frequency response. The results of both methods were compared. The trajectories of formants are more stable, when they are obtained using a second approach. Ill. 7, bibl. 8, tabl. 1 (in English; abstracts in English, Russian and Lithuanian).

Д. Балбонас, Г. Даунис. Моделирование изменений спектра гласных используя акустическую трубу // Электроника и электротехника. – Каунас: Технология, 2009. – № 8(96). – С. 85–88.

Вокальный тракт может быть описан с использованием сериально связанных электрических цепей Т-типа. Его частотная характеристика была получена двумя способами: моделированием сигнала с последующим расчетом спектра и вычислением частотной характеристики системы сериально связанных четырехпольников. В первом случае, математическая модель акустической трубы была построена с использованием методам переменых состояния, где переменные состояния - напряжения на конденсаторах (эквиваленты звукового давления) и токи через индуктивности (эквиваленты скорости объема воздуха). Получена система линейных дифференциальных уравнений была интегрирована косвенным методом. Во втором случае „а" параметры четырехпольников были использованы для получения амплитудно-частотной характеристики всей системы. Во время произглашения гласных площади сечений вокального тракта изменяются с течением времени. В обоих случаях траектории формантов были получены из частотных характеристик. Результаты обоих методов были сравнены. Получено, что траектории формантов получены с помощью второго подхода, более стабильны. Ил. 7, библ. 8, табл. 1 (на английском языке; рефераты на английском, русском и литовском яз.).

D. Balbonas, G. Daunys. Balsių spektro kitimo modeliavimas naudojant akustinį vamzdį // Elektronika ir elektrotechnika. – Kaunas: Technologija, 2009. – Nr. 8(96). – P. 85–88.

Kalbos traktas gali būti atvaizduotas, naudojant nuosekliai sujungtas T tipo elektrines grandines. Dažninės charakteristikos buvo modeliuojamos dviem būdais: modeliuojant signalą bei skaičiuojant jo spektrą ir skaičiuojant nuosekliai sujungtų keturpolių dažninę charakteristiką. Pirmuoju atveju matematinis akustinio vamzdžio modelis sudarytas taikant būsenos kintamųjų metodą, kai būsenos kintamieji yra kondensatorių įtampos (atitinka garso slėgius) ir induktyvumų srovės (atitinka oro tūrio greičius). Gauta tiesinių diferencialinių lygčių sistema integruojama netiesioginiu metodu. Antruoju atveju skaičiuojami visų keturpolių „a" parametrai ir iš jų gaunama visos sistemos dažninė charakteristika. Tariant balsius, kinta kalbos trakto skerspjūviai, todėl elementų parametrai laikui bėgant kinta. Abiem metodais gavus kalbos trakto dažnines charakteristikas, buvo nustatyti formantai. Abiem metodais gauti rezultatai palyginti tarpusavyje. Modeliuojant antruoju metodu, gaunamos pastovesnės formantų kitimo kreivės. Il. 7, bibl. 8, lent. 1 (anglų kalba; santraukos anglų, rusų ir lietuvių k.).