ELECTRONICS AND ELECTRICAL ENGINEERING

ISSN 1392 – 1215 2007. No. 7(79)

ELEKTRONIKA IR ELEKTROTECHNIKA

AUTOMATION, ROBOTICS

T125

AUTOMATIZAVIMAS, ROBOTECHNIKA

Movement of Formants of Vowels in Lithuanian Language

D. Balbonas, G. Daunys

Department of Electronics, Siauliai University,
Vilniaus str. 141, LT-76353 Šiauliai, Lithuania, phone: +370 41 595830; e-mail: dainius@tf.su.lt

Introduction

Using of voice technologies is challenging approach in human computer interaction. Fine structure of speech must be revealed for natural voice synthesis and robust speech recognition. Spectral analysis is common accepted research method for speech analysis. Spectral formants are very informative variables of voiced speech sounds. It is possible that their movement analysis could help to build more accurate model of speech.

L. Kaukienė [1] analyzed the spectra of unstressed vowels in Standard Lithuanian, their acoustic and articulations characteristics. The unstressed vowels were compared with the stressed ones. She found that low rise vowels "a", "e" have the biggest qualitative reduction.

R Kliukienė [2] investigated regressive influence of palatalized consonants on the spectrum of the short vowels in Standard Lithuanian. She found that palatalized consonants increase frequency of second formant.

D. Balbonas, G. Daunys [3] analyzed the formant movement of vowel "a" in Lithuanian language. They found that formants movement in vowel "a" depends upon stress. Sound duration and variation of intensity are important parameters, which allow to model pronunciation of sound "a".

J. M. Hillenbrand, T. M. Nearey [4] examines the role of formant frequency movements in vowel recognition. They found no evidence of a simple inverse relationship between F0 and vowel intelligibility.

Also they found that spectral change patterns play a secondary but quite important role in the recognition of vowel quality. The relative importance of formant frequency change varies considerably from one vowel to the next. The relationships between spectral change patterns and vowel identity are guaranteed to be more complex when the consonant environment preceding and following the vowel is allowed to vary.

Authors tried to find out how to explain formant movement in Lithuanian vowels "a", "e", "ė", "i", "u" and "o". The first idea was that so happens because of reduction. But our investigation revealed that for first two formants most influence have stress position in the word. Also authors have attempted to find the generalized frequencies of the first and second formants of all Lithuanian vowels,

Method

The authors recorded the speech of three speakers; two males and one female. The speech was recorded in quiet environment using computer microphone. For each speaker was prepared text with 635 words (774 syllables). The whole text have 93 words (152 syllables) with sound "a", 112 words (117 syllables) with sound "e", 73 words (79 syllables) with sound "ė", 137 words (196 syllables) with sound "i", 97 words (100 syllables) with sound "o" and 123 words (130 syllables) with sound "u". The words selected for speech was elected from 1200 frequently used words in Lithuanian language. The speaker's speech files with different sounds were divided into separate sound files and then using Praat software sounds a, e, ė, i, o, u and transition between these sounds and other sounds in the concrete word were marked [5]. Marking was made manually. Information about marking was saved in Text Grid files (Praat extension). Each word has its own Text Grid file that contains information about location of syllables and empty intervals in the word. Also file contains information about the names of syllables, the length of word and all marked intervals in the word.

When using specialized Praat script the points of F0, F1, F2, F3 F4 F5 formants at every 10ms were extracted from wave file to Excel data sheet, for each word. In the end we had 2322 Excel files with formant points and 2322 Text Grid files. Text Grid and Excel files had used in Matlab program as a data for artificial neural network.

The artificial neural networks (ANN) approach was used to find approximation of formants frequency on factors. All data was divided into three groups for training, validation and testing. So the earlier stopping was achieved. The attempts were done to find smallest set of factors which best describe formants frequencies.

Results and discussion.

The analysis of formants frequencies distribution of vowels for three subjects was done. The results for F1 and F2 are presented in Table 1. The data in the table was collected from formants frequencies distribution in each vowel of each subject. For example Fig. 1 shows all formants frequencies distribution of vowel "e" for subject V1.

Table 1. Distribution of formant F1 and F2 frequencies

Formant	F1			F2		
Subject/ vowels	V1	V2	M1	V1	V2	M1
a	490-1090	300-750	590-1070	1050-1620	1050-1650	1070-1770
e	410-1000	350-800	460-1010	1300-2060	1300-1700	1130-2260
i	240-570	250-500	240-600	1740-2410	1150-2400	1740-2640
u	310-750	250-450	230-640	690-2110	660-1600	720-1620
o	470-770	320-610	430-910	710-1370	860-1350	780-1300
ė	330-580	300-660	340-820	1780-2310	1340-2160	1350-2540

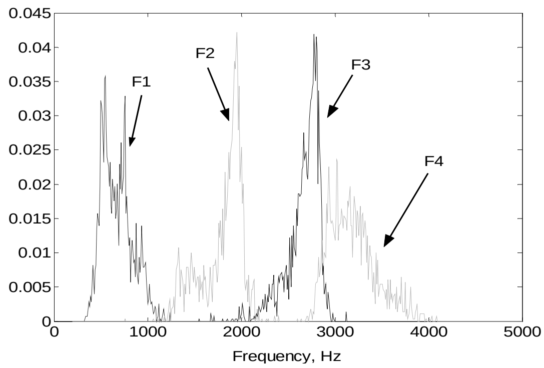

Fig. 1. F1, F2, F3, F4 formants frequencies distribution of vowel "e" of V1 subject

Artificial neural network was used to find out how formants movement dependence from stress. The best results were obtained, when polynomial coefficients of intensity regression versus time were used together with sound duration and time moment of sample.

Also the distribution of all formants frequency errors after approximation with ANN for all vowels and subjects were done. Results of distribution of F1 and F2 formants frequency errors are presented in Table 2.

Table 2. Distribution of formants F1, F2 frequency errors

Formant	F1			F2		
Subject/ vowels	V1	V2	M1	V1	V2	M1
a	±100	±40	±70	±80	±70	±100
e	±70	±40	±60	±90	±40	±200
i	±50	±40	±60	±90	±150	±150
u	±70	±30	±70	±110	±90	±150
o	±50	±30	±90	±110	±40	±100
ė	±30	±40	±60	±50	±70	±150

The distribution of remaining errors for vowel "i" (F1 formant, V1 subject) is plotted in Fig. 2.

The distribution is close to Gaussian distribution. The interval, where the errors most frequently occur is ±50 Hz (95%-97% of samples are ±50 Hz range) for F1 and about ±100 Hz for F2. For F2 formant the bigger error rate was found for M1 subject.

The simulation results of F1 formant (all vowels) movement for V2 subject versus time, when other parameters are constant (they have their mean values of stress group) are shown in Fig. 3–8. The four cases of stressed syllables are used (0 – non-stressed, 1 – grave, 2 – acute accent, 3 – circumflex).

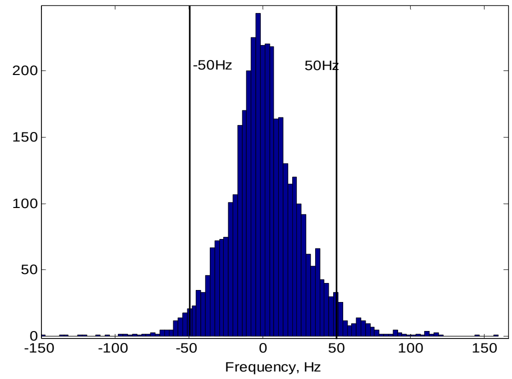

Fig. 2. The distribution of remaining errors for vowel "i" (F1 formant, V1 subject)

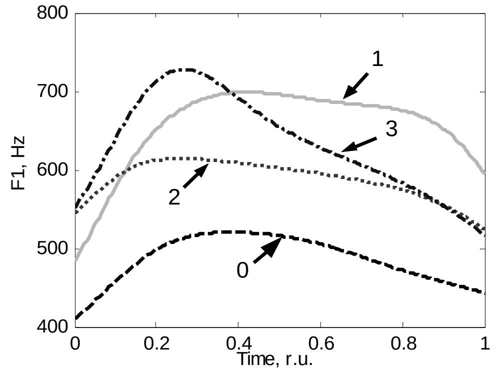

Fig. 3. Formant F1 movement during utterance of sound "a" versus normalized time in case of differently stressed syllable

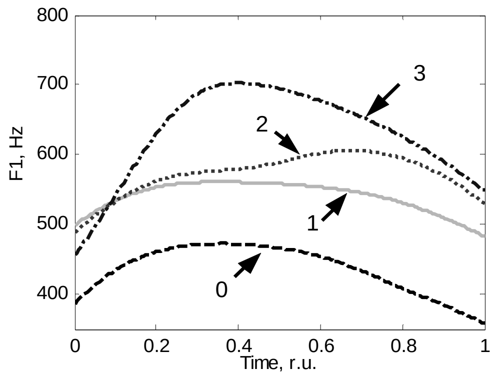

Fig. 4. Formant F1 movement during utterance of sound "e" versus normalized time in case of differently stressed syllable

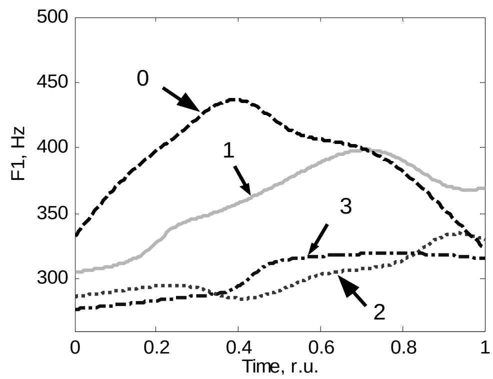

Fig. 5. Formant F1 movement during utterance of sound "i" versus normalized time in case of differently stressed syllable

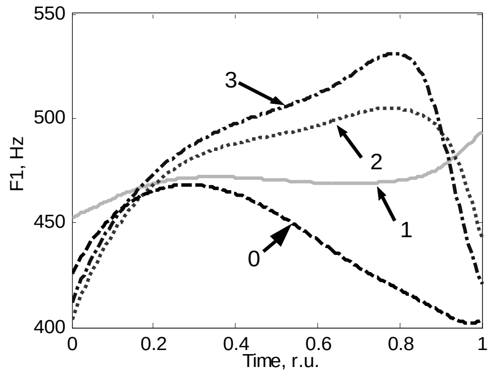

Fig. 6. Formant F1 movement during utterance of sound “o” versus normalized time in case of differently stressed syllable

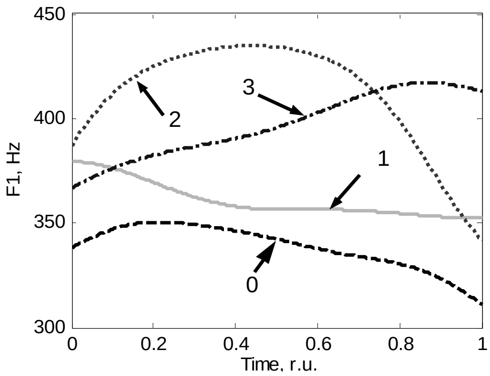

Fig. 7. Formant F1 movement during utterance of sound “u” versus normalized time in case of differently stressed syllable

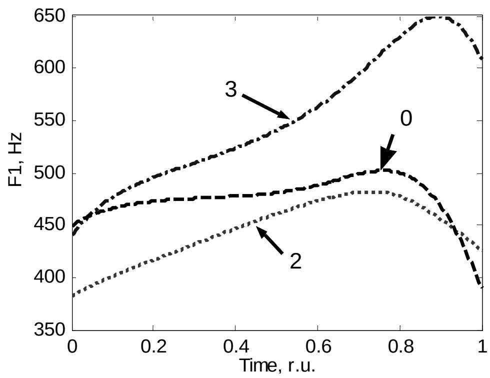

Fig. 8. Formant F1 movement during utterance of sound “ė“ versus normalized time in case of differently stressed syllable

The authors analyzed the data in Fig. 3–8 and have noticed that there is no difference between vowels “a” and “e” in F1. Exception is for the grave stress because the frequency of F1 in vowel “a” is higher than in vowel “e”. Also analyzes of figures shows that F1 frequency is the lowest for non-stressed vowels except vowel “i”. F1 frequency of vowel “i” for non-stressed is the highest. The F1 frequency, for 1, 2 and 3 case stress, is higher in “e” and “a” than in “ė”, “i”, “o”, “u”. Fig. 8 shows that vowel “ė” do not have the grave stress

General information about F1 formant for V2 subject in Lithuanian vowels is shown in Table 3. Fig. 3–8 and Table 1 show us that F1 formant frequency and moving way depends on stress. Table 2 shows general information about F2 formant for V2 subject in Lithuanian vowels.

Tables 3 and 4 can be useful for sound synthesis or recognition, for example there is a need to synthesized sound non-stressed “a”.

Table 3. F1 formant grouping according stress case for V2 subject

Stress / case	The frequency range, Hz			Form of formant movement*		
	350-550	300-350	325-450	mal	mar	c
0	a, e, ė, o	u	i	a, e, i, o	ė	u
	500-700	450-550	300-400	mal	mar	c
1	a	e, o	i, u	i	a, e	u, o
	500-625	350-500	300-350	mal	mar	su
2	a, e	ė, u, o	i	e, ė, o	a, u	i
	425-750	375-425	275-325	mal	mar	su
3	a, e, ė, o	u	i	a, e	ė, o	i, u

*c – constant, mal – Maximum on the left, mar – Maximum on the right, su – small up

Table 4. F2 formant grouping according stress case for V2 subject

Stress / case	The frequency range			Form of formant movement*			
	1200-1500	1425-1725	1050-1300	d	sd	u	ld
0	a, ė, u	e , i	o	a, ė	e	u, o	i
	1675-1800	1125-1550	675-950	lu	c	mar	u
1	i	a, u, e	o	u, o	e	a	i
	1625-2400	800-1225	1550-1575	c	ld	mil	d
2	ė, i	o, u, a	e	a, e	ė, i	o	u
	1350-1575	1550-2500	850-1225	sd	u	ld	mac
3	a, e	ė, i	o, u	e, u	a, o	i	ė

*c – constant, mar – maximum on the right, d – down, u – up, sd - small down, ld – large down, mil – minimum on the left, mac – maximum on the centre, lu – large up

The frequency range for F1 is 400-525Hz the same frequency range can have: non-stressed “a”, “e”, “ė”, “i”, acute accent “ė”, “u”, sound “o” in all situation. So this frequency range can be 10 sounds, but the same form of formant have only 6 sounds: non stressed “a”, “e”, “i”, “o” and acute accent “ė”, “o”. In next step we can add information from table 4. The frequency range for F2 is 1250-1500Hz the same frequency range can have: non-stressed “a”, “ė”, “i”, “o”, “u”, grave “i”, “u” and circumflex “a”. The same form has only non-stressed “a”, “ė”, “i”. From both tables only non-stressed “a” and “i” have the same frequency range and formant form in F1 and F2. Really F2 formant form for sound “a” and “i” are not very similar, sure they both are d (down – frequency descending) type, but differences is between descending style - sound “a” is convex styled, sound “i” is concave styled. So the question is it possible, using frequency range and formants style, to separate different sounds?

The authors calculated the generalized frequencies of F1, F2 formants for each subject. The centre frequencies of Lithuanian vowels are shown in Fig. 9–11.

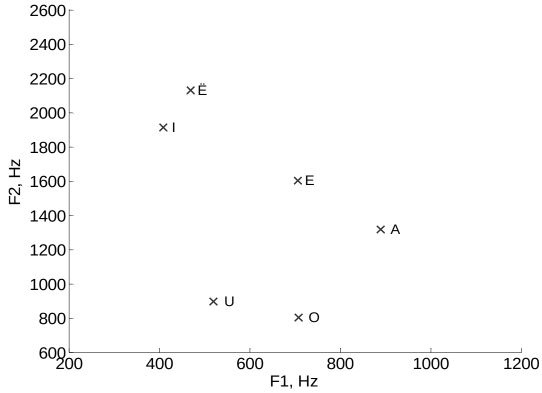

Fig. 9. Generalized frequencies of F1 and F2 for V1 subject

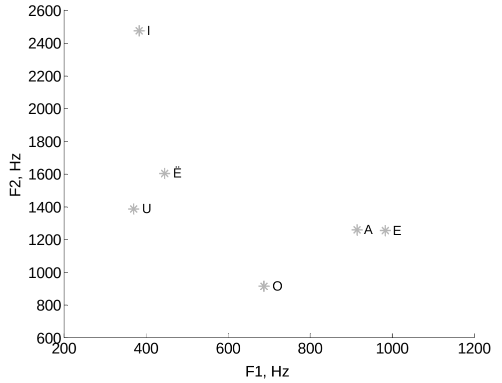

Fig. 10. Generalized frequencies of F1 and F2 for M1 subject

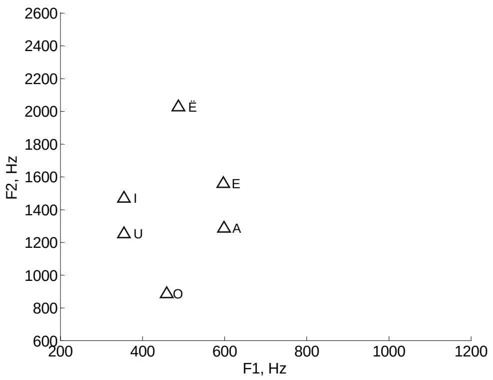

Fig. 11. Generalized frequencies of F1 and F2 for V2 subject

From Fig. 9–11 we can see that several F1 and F2 formants predicted frequencies vary from subject to subject.

Conclusions

As in our previous study [3] for sound "a", we obtained that the formant movement of Lithuanian vowels depends upon stress. The good approximation of experimental results of F1 and F2 by ANN was obtained, when the polynomial coefficients of intensity approximation upon time were used as input parameters together with sound duration and current moment time.

The map of generalized frequencies of formant F1 and F2 strongly depends of subject. It is issue for using them in speech recognition and future investigations in this field are needed.

References

1. **Kaukėnienė L.** The spectra of unstressed vowels in standard Lithuanian // Kalbotyra. – LI (1). – 2002.
2. **Kliukienė R.** Regressive influence of palatalized consonants on the spectrum of the short vowels in Standard Lithuanian // Kalbotyra. – LI(1). – 2002. – P. 73–78.
3. **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.
4. **Hillenbrand J. M., Nearey T. N.** Identification of resynthesized /hVd/ syllables: Effects of formant contour // Journal of the Acoustical Society of America. – 1999. – No. 105. – P. 3509–3523.
5. **Praat** project homepage. Reachable by Internet: http://www.praat.org/.

Submitted for publication 2007 03 28

D. Balbonas, G. Daunys. Movement of Formants of Vowels in Lithuanian Language // Electronics and Electrical Engineering. – Kaunas: Technologija, 2007. – No. 7(79). – P. 15–18.

The analysis of spectral formant movement of Lithuanian vowels was carried out. More frequently used words with different surrounding of sounds *"a", "e", "ė"," i", "o", "u"* were selected for investigation. Praat program was used for words annotation into phonemes and obtaining values of formants. The movement of formants was approximated by other sound parameters, as duration, time moment from sound pronunciation beginning, intensity and its polynomial approximation coefficients. Approximation was carried out, using artificial neural network. The result was more narrow remaining errors distribution. The sounds were divided into groups by their stress, because parameters of approximation have less spread in these groups. The formant movement dependencies versus time were obtained. The generalized frequencies for formants F1 and F2 were calculated, when time is in the middle of the sound and other parameters have zero values. Ill. 11, bibl. 5 (in English; summaries in English, Russian and Lithuanian).

Д. Бальбонас, Г. Даунис. Движение формант гласных в литовском языке // Электроника и электротехника. – Каунас: Технология, 2007. – № 7(79). – С. 15–18.

Произведен анализ движения формант гласных литовского языка был. Отобраны более часто используемые слова, в которых в различном контексте находятся звуки "a", "e", "ė", "i", "o", "u". Для аннотации звуковых файлов была использована Praat программа. Значения формант тоже были рассчитаны с помощью Praat программы. Произведена аппроксимация значений формант от других параметров звука, таких как длительность, момент времени, интенсивность и коэффициенты её полиномной аппроксимации, было произведено. Для аппроксимации была применена искусственная нейронная сеть. Аппроксимация позволила в несколько раз уменьшить разброс значений. Звуки были сгруппированы по их ударению, так как разброс параметров аппроксимации в этих группах несколько раз меньше. Найдены зависимости движения формант во времени. Обобщенные значения формант найдены в середине временного интервала при других параметрах, равных нулю. Ил. 11, библ. 5 (на английском языке; рефераты на английском языке, русском и литовском яз.).

D. Balbonas, G. Daunys. Lietuvių kalbos balsių formantų kitimas // Elektronika ir elektrotechnika. – Kaunas: Technologija, 2007. – Nr. 7(79). – P. 15–18.

Atlikta lietuvių kalbos balsių spektro formantų kitimo analizė. Tyrimui buvo pasirinkti dažniausiai lietuvių kalboje pasitaikantys žodžiai, kuriuose garsai *„a", „ e", „ ė", „i", „o", „u"* yra apsupti įvairių garsų. Praat programa žodžiuose buvo pažymėti garsų intervalai ir nustatyta, kaip juose keičiasi keturi pirmieji formantai. Atliktas formantų kitimo priklausomai nuo kitų garsų parametrų, tokių kaip trukmė, laiko momentas nuo balsio tarimo pradžios, intensyvumas bei jo skleidimo ketvirtos eilės polinomu koeficientai, aproksimavimas. Aproksimuota dirbtinių neuronų tinklu. Toks aproksimavimas kelis kartus sumažina liekamąją sklaidą. Garsai buvo suklasifikuoti pagal jų kirčiavimą, kadangi šiose grupėse aproksimavimo parametrų sklaida yra pastebimai mažesnė. Surastos formantų kitimo bėgant laikui priklausomybės. Taip pat apskaičiuoti apibendrinti formantams F1 ir F2 dažniai, imant laiką intervalo viduryje, o kitus parametrus prilyginant nuliui. Il. 11, bibl. 5 (anglų kalba; santraukos anglų, rusų ir lietuvių k.).