Liet. mat. rink. LMD darbai, **50**, 2009, 391–396

Application of the MUSIC method for estimation of the signal fundamental frequency

Virginija ŠIMONYTĖ[1], Gražina PYŽ[2], Vytautas SLIVINSKAS[1,3]

[1] Vilnius Pedagogcal University, Faculty of Mathematics and Informatics
Studentų 39, LT-08106 Vilnius, Lithuania

[2] Insitute of Mathematics ana Informatics
Akademijos 4, LT-08663 Vilnius, Lithuania

[3] Vilnius Gedimino Technical University, Faculty of Fundamental Sciences
Saulėtekio 11, LT-10223 Vilnius, Lithuania

e-mail: virgin0525@gmail.com; grazula@takas.lt; vytas@astera.lt

Abstract. The goal of the paper is to use the MUSIC method to estimate the fundamental frequency of signals, and compare the results with the ones obtained by the DFT method. Real speech signals are considered in the paper. Lithuanian vowel sound fundamental frequency is estimated by the MUSIC and DFT methods.

Keywords: MUSIC method, fundamental frequency, discrete Fourier transform (DFT), Lithuanian speech signals.

1. Introduction

Estimation of a fundamental frequency is very important in many fields of speech signal processing such as speech coding, speech synthesis, speech and speaker recognition [1,9]. The speech signal fundamental frequency is an essential feature of human voice [2]. What we hear as a single sound when someone is speaking (for example, pronouncing /a/) is really the fundamental frequency plus a series of harmonics. The fundamental frequency is determined by the number of times the vocal folds vibrate in one second, and measured in cycles per second [cps], or Hertz [Hz]. The harmonics are multiples of the fundamental frequency. Thus if the fundamental frequency is 100 Hz, the harmonics are 200 Hz, 300 Hz, 400 Hz, etc. Mathematically, if $y(t)$ is a sound signal, then we can use the following model:

$$y(t) = \sum_{k=1}^{p} a_k \sin(2\pi f_0 kt + \varphi_k) + e(t), \tag{1}$$

where $a_k \in R$, $\varphi_k \in [-\pi, \pi]$, $\{e(t)\}$ is white Gaussian noise; the lowest frequency f_0 is called the fundamental frequency, and other frequencies $f_k = kf_0$ $(k = 2, \ldots, p)$ are called harmonics. The fundamental frequency is also called the first harmonic. We normally don't hear the harmonics as separate tones, they, however, exist in the sound and add a lot of richness to the sound. Without them a voice would sound uninteresting

and synthetic [16]. Often the sinusoid of the frequency $f_k = kf_0$ is itself called the kth harmonic of the signal $y(t)$.

Much efforts are given in Lithuania for developing digital technologies of Lithuanian speech processing [3–7,11]. Lithuanian speech synthesis is one of the tasks of Lithuanian speech digital processing. In order to solve the problem of Lithuanian speech synthesis, it is necessary to develop mathematical models for Lithuanian speech sounds. Developing of the vowel models is a part of this problem. One of the main vowel models is a model of harmonically related sinusoids. The main parameter of this model is the fundamental frequency. In order to get good quality of a synthesised sound, one needs to estimate this frequency as accurately as possible. The DFT method is usually used to estimate this frequency. This method gives good results when the observed signal is sufficiently long. For shorter signals, performance of this method is not satisfactory. Thus alternative methods have to be used. One of such algorithms is the so-called MUSIC method. This method is used widely in the mobile communications field. In [10] T. Murakami and Y. Ishida applied the MUSIC method for the analysis of speech signals. They used this method for the fundamental frequency estimation of Japanese female and male vowels /a/, /e/, /i/, /o/, /u/, and illustrated that their method based on the MUSIC method is superior to the conventional cepstral method for estimating the fundamental frequency.

The goal of this paper is to apply the MUSIC method for estimation of the fundamental frequency of the main Lithuanian vowels. This paper is organized as follows. The MUSIC algorithm is reviewed in Section 2. We present comparison of the results obtained by the MUSIC method and the conventional DFT method in Section 3. Section 4 contains the conclusions.

2. MUSIC method

Consider the following model:

$$y_n = \sum_{l=1}^{p} c_l \exp(j w_l n) + e_n \quad (n = 1, \ldots, N), \tag{2}$$

where $c_l \in C$, $\{e_n\}$ is white noise. Let M be some integer greater than p. Define

$$\begin{aligned} y(t) &= [y_t, \ldots, y_{t+M-1}]^T, \\ x(t) &= \left[c_1 e^{j w_1 t}, \ldots, c_p e^{j w_p t}\right]^T, \\ e(t) &= [e_1, \ldots, e_{t+M-1}]^T, \end{aligned} \tag{3}$$

where $t = 1, \ldots, N - M + 1$. Define also

$$\begin{aligned} a(w) &= \left[1, e^{jw}, \ldots, e^{j(M-1)w}\right]^T, \\ \theta &= [w_1, \ldots, w_p]^T, \\ A(\theta) &= \left[a(w_1), \ldots, a(w_p)\right]. \end{aligned} \tag{4}$$

We can now write (2) as

$$y(t) = A(\theta)x(t) + e(t) \quad (t = 1, 2, \ldots, K = N - M + 1). \tag{5}$$

The MUSIC method [12,13,15] was developed in 1979 by American scientist R. Schmidt. The acronym MUSIC stands for MUltiple SIgnal Classification. This method deals with estimation of parameters of (5) model.

The covariance matrix $R = Ey(t)y^H(t)$ of the vector $y(t)$ is given by [14]

$$R = A(\theta)PA^H(\theta) + \sigma^2 I_{M\times M}, \tag{6}$$

where σ^2 is as in $Ee(t)e^H(t) = \sigma^2 I_{M\times M}$, and $P = Ex(t)x^H(t)$.

Denote by $\lambda_1 > \lambda_2 > \ldots > \lambda_M$ the eigenvalues of the matrix R. Since rank $(APA^H) = p$ [14], then

$$\lambda_k > \sigma^2 \quad (k = 1, \ldots, p) \quad \text{and} \quad \lambda_k = \sigma^2 \quad (k = p+1, \ldots, M). \tag{7}$$

Let $s_1, s_2, \ldots, s_p$ be the unit-norm eigenvectors corresponding to the first p largest eigenvalues $\lambda_1, \lambda_2, \ldots \lambda_p$, and $g_1, g_2, \ldots, g_{M-p}$ – the unit-norm eigenvectors corresponding to the last $M - p$ smallest eigenvalues $\lambda_{p+1}, \lambda_{p+2}, \ldots, \lambda_M$. Denote by S an $M \times p$ matrix whose columns are the vectors $s_1, s_2, \ldots, s_p$, and by G an $M \times (M - p)$ matrix whose columns are the vectors $g_1, g_2, \ldots, g_{M-p}$, i.e.,

$$S = [s_1, \ldots, s_p], \quad G = [g_1, \ldots, g_{M-p}]. \tag{8}$$

It is shown in [14] that the true parameter values $\{w_1, \ldots, w_p\}$ are the only solutions of the following equation: $a^H(w)GG^H a(w) = 0$.

In practice, we use an estimate

$$\hat{R} = \frac{1}{M}\sum_{t=1}^{M} y(t)y^H(t) \tag{9}$$

of the true covariance matrix R. Denote by $\hat{s}_1, \ldots, \hat{s}_p, \hat{g}_1, \ldots, \hat{g}_{M-p}$ the unit-norm eigenvectors of $\hat{R}$ arranged in the descending order of the corresponding eigenvalues, and by $\hat{S}$ and $\hat{G}$ – the matrices made of $\{\hat{s}_1, \ldots, \hat{s}_p\}$ and $\{\hat{g}_1, \ldots, \hat{g}_{M-p}\}$. Define the MUSIC spectral function as follows:

$$\hat{P}_{MU}\left(e^{jw}\right) = \frac{1}{a^H(e^{jw})\hat{G}\hat{G}^H a(e^{jw})}. \tag{10}$$

The estimates of $\{w_1, \ldots, w_p\}$ are obtained by maximizing $\hat{P}_{MU}(e^{jw})$. This procedure is done by evaluating it at the points of a fine grid.

3. Fundamental frequency estimation using the MUSIC method

We consider real data in this section. This data is samples of natural sounds. The sounds were recorded using a microphone and the "Sound Record" program. The sound recording parameters were as follows: the sampling frequency equal to 48 kHz,

and the signal quantization accuracy – 16 bits. This frequency corresponds to the sampling interval of 21 μs. The experiments were carried out using our programs developed in MATLAB. We applied the MUSIC method and DFT (Discrete Fourier Transform) method for Lithuanian female vowels /a/, /i/, /o/, /u/. The MUSIC spectrum was calculated using the formula

$$\text{MUSIC}(f) = 10 \cdot \lg\left(\left|\hat{P}_{MU}\left(e^{j2\pi f}\right)\right|\right), \tag{11}$$

where $\hat{P}_{MU}(e^{j2\pi f})$ is defined by (10).

For each of the vowels mentioned above, we considered 80 records of length 1024 points. For each of these records, we calculated the spectra and estimates of the fundamental frequency by the MUSIC method and DFT, and obtained their mean $E(\hat{f}_0)$ and standard deviation $\sigma(\hat{f}_0)$. The results are shown in Fig. 1 and Table 1.

We see from Table 1 that the vowel /i/ has the highest fundamental frequency. The vowel /a/ has the second highest fundamental frequency. This fact can be observed in the estimation results of both methods. The lowest fundamental frequency is in the vowel /u/ (the MUSIC result) or in the vowel /o/ (the DFT result). The difference between the estimated frequencies of the vowels /o/ and /u/, however, is small – about 1 Hz (0.85 Hz for the MUSIC, and 1.17 Hz for the DFT). The smallest standard deviation 2.36 Hz was obtained by the MUSIC method for the vowel /i/, and the largest – 5.59 Hz – by the DFT method for the vowel /a/. It is easy to notice that the DFT standard deviation values are higher than those of the MUSIC method for all vowels.

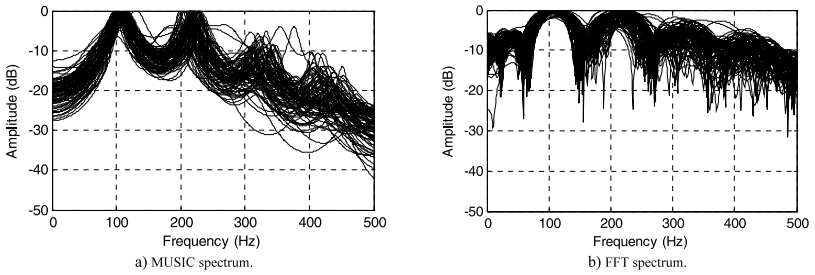

Fig. 1. Spectrum estimates for a Lithuanian female vowel /u/ for 80 speech signal realisations.

Table 1. The mean and standard deviation of the fundamental frequency estimates obtained by the MUSIC method and DFT method

Characteristics	Vowel			
	/a/	/i/	/o/	/u/
The mean $E(\hat{f}_0)_{MUSIC}$	112.00	114.75	108.20	107.35
The mean $E(\hat{f}_0)_{DFT}$	124.80	127.15	118.95	120.12
The standard deviation $\sigma(\hat{f}_0)_{MUSIC}$	3.55	2.36	4.13	4.62
The standard deviation $\sigma(\hat{f}_0)_{DFT}$	5.59	4.18	4.18	5.07

Table 2. The relative approximation error of the vowel signals by the sum of ten harmonics using the fundamental frequency estimates obtained by the MUSIC method and DFT method

Error	Vowel			
	/a/	/i/	/o/	/u/
err_{MUSIC}	37.32%	44.74 %	22.99%	8.86 %
err_{DFT}	78.75%	54.70%	68.33 %	37.94 %

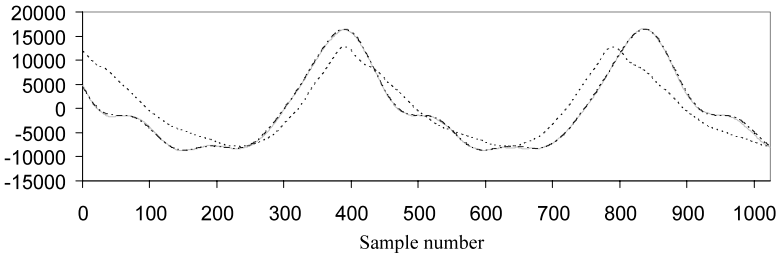

Fig. 2. The true and estimated speech signal of the vowel /u/ (solid line – the true speech signal, dotted line – the estimated signal (DFT method), dash-dotted line – the estimated signal (MUSIC method)).

Since the DFT and MUSIC methods give different estimates of the fundamental frequency, we have to check which estimate describes the real situation more accurately. For each vowel, we used a model of the sum of ten harmonics where the first harmonic frequency was the fundamental frequency estimate (the DFT or MUSIC). The parameters of the harmonics were estimated by a usual linear least squares method. The relative estimation errors are shown in Table 2.

We see from Table 2 that the errors obtained with the MUSIC fundamental frequency estimate are smaller than those obtained with the DFT fundamental frequency estimate. The fact that the errors are rather large can be attributed to complexity of the real sound signals. Harmonics of such signals are time variant, and it is very difficult to describe the signal with time invariant frequency models.

The true signal of the vowel /u/ and its estimates obtained by the DFT and MUSIC methods are shown in Fig. 3. The signal estimates are obtained using a model of the sum of 10 harmonics. One can see that the MUSIC estimate almost coincides with the true signal.

4. Conclusions

Estimation of the fundamental frequency is very important for synthesis of Lithuanian speech vowels since this frequency is the main parameter of the vowel models. It shows a frequency at which impulses must be given to the input of the synthesizer forming filter.

The model (1) can be used in vowel recognition. A vector made of the fundamental frequency and amplitudes of the first 10–20 harmonics can be taken as a recognition vector. The harmonic amplitudes are obtained by a simple least squares fit.

Our investigation has shown that the estimates of the fundamental frequency obtained by the MUSIC method are less scattered around their average if compared with the ones obtained by the DFT method.

Approximation of a real sound signal by a sum of the first 10 harmonics with the fundamental frequency obtained by the MUSIC and DFT methods gave a smaller error in the case of the MUSIC method.

Acknowledgement. The authors are grateful to the anonymous reviewer whose comments helped to improve the quality of this paper.

References

1. A. de Cheveigné, H. Kawahara. YIN, a fundamental frequency estimator for speech and music. *The Journal of the Acoustical Society of America*, 111(4):1917–1930, 2002.
2. W. Hess. *Pitch Determination of Speech Signals*. Springer-Verlag, New York, 1983.
3. P. Kasparaitis. *Text-to-Speech Synthesis of Lithuanian Language*. Doctoral dissertation, Vilnius University, Vilnius, 2001 (in Lithuanian).
4. P. Kasparaitis. Transcribing of the Lithuanian text using formal rules. *Informatica*, 10(4):367–376, 1999.
5. P. Kasparaitis. Automatic stressing of the Lithuanian text on the basis of a dictionary. *Informatica*, 11(1):19–40, 2000.
6. A. Lipeika, J. Lipeikienė. Speaker identification. *Informatica*, 4(1–2):45-56, 1993.
7. A. Lipeika, J. Lipeikienė. Language engineering in Lithuania. *Informatica*, 9(4):449–456, 1998.
8. J. Mikelionienė. Lithuanian language processing using digital technologies. *Tiltai*, 2(31):91–96, 2005.
9. Z. Milivojevic, M. Mirkovic, S. Milivojevic. An estimate of fundamental frequency using PCC interpolation – comparative analysis. *Information Technology and Control*, 35(2):131–136, 2006.
10. T. Murakami, Y. Ishida. Fundamental frequency estimation of speech signals using MUSIC algorithm. *Acoust. Sci. Technol.*, 22(4):293–297, 2001.
`http://www.jstage.jst.go.jp/article/ast/22/4/293/_pdf`.
11. T. Ringys, V. Slivinskas. Formant modelling of natural sounding of Lithuanian vowels. In: *4th international conference "Electrical and Control Technologies ECT-2009"*, 7–8 May, 2009, Kaunas, Lithuania, 5–8, 2009 (in Lithuanian).
12. R.O. Schmidt. Multiple emitter location and signal parameter estimation. *IEEE Trans. Antennas Propagation*, 34(3):276–280, 1986.
13. P. Stoica, R. Moses. *Introduction to Spectral Analysis*. Englewood Cliffs, Prentice-Hall, 1997.
14. P. Stoica, A. Nehorai. MUSIC, maximum likelihood, and Cramer-Rao bound. *IEEE Trans. Acoustics, Speech, and Signal Processing*, 37(5):720–741, 1989.
15. C.W. Therrien. *Discrete Random Signals and Statistical Signal Processing*. Englewood Cliffs, Prentice-Hall,
16. `http://homepage.ntu.edu.tw/~karchung/phonetics%20II%20page%20eight.htm`.

REZIUMĖ

V. Šimonytė, G. Pyž, V. Slivinskas. MUSIC metodo taikymas signalų fundamentalaus dažnio nustatymui

Darbo tikslas yra panaudoti MUSIC metodą signalų fundamentalaus dažnio įvertinimui, palyginti gautus rezultatus su DFT metodu. Straipsnyje nagrinėjami tiek modeliniai, tiek realūs signalai. Naudojant MUSIC metodą, gauti lietuviškų balsių garsų fundamentalaus dažnio įverčiai, kurie yra palyginti su DFT metodu gautais įverčiais.

Raktiniai žodžiai: MUSIC metodas, fundamentalus dažnis, diskrečioji Furjė transformacija (DFT), lietuvių kalbos signalai.