Cepstral Vocal Tract Modelling for Text-To-Speech Synthesis

Dr. Jafar Al-Kheir* **Dr. Zdenek Smékal****

Abstract

In this paper we describe a cepstral model of the vocal tract which models both formants and anti-formants. The investigated model is more precise compared to the linear prediction model, which models only the formants of the vocal tract. The exponential function is used for the inverse transformation. However, it is difficult to implement this function on a digital signal processor. To solve this issue we use a continued fraction expansion to approximate the exponential function. The transfer function that approximates the exponential function is realized by using the Infinite Impulse Response (IIR) digital filter, in which branches type Finite Impulse Response (FIR) digital filters are included. The coefficients of the FIR digital filters are just the coefficients of the real speech cepstrum. The state-space difference equations are proposed and implemented on a DSP56300 fixed-point digital signal processor (Motorola). Finally, the results of the digital signal processor implementation for chosen vowels and consonants are evaluated.

Keywords: Real Speech Cepstrum, Vocal Tract Model, Digital Signal Processor, Text-To-Speech Synthesis.

* Tishreen University- Lattakia,
** Department of Telecommunications- Brno University of Technology- Purkynova

INTRODUCTION

The parametric method is one of the speech production methods used in text-to-speech (TTS) synthesis Anexcitation signal excites the vocal-tract model with time-varying parameters. A new state-space cepstral vocal-tract model is described, which approximates both the formants and the anti-formants of the model frequency response for voiced and unvoiced speech sounds. The proposed model differs from the currently used Linear Predictive Coding (LPC) model, which approximates only the formants alone [1]. Unlike methods of the type of Pitch Synchronous OverLap and Add (PSOLA), the investigated method is convenient for prosody modelling and requires less memory capacity. The cepstral speech synthesis starts from the cepstral coefficients obtained by analysing the speech signal. Generally, there are two basic architectures used for design the digital signal processor. The “Von Neumann” architecture is composed of a single memory and a single bus for transferring data into and out of the central processing unit (CPU). Because of that, multiplying two numbers requires at least three clock cycles, one to transfer each of the two numbers and the result over the bus from the memory to the CPU. Unlike the “Von Neumann” architecture, the “Harvard” architecture insisted on two separate memories with separate buses for data and program instructions. Because of that, program instructions and data canbe fetched at the same time, which results in improving the speed over the single bus design. Most nowadays DSPs use this dual bus architecture.
Digital signal processing can be classified into two categories - fixed point and floating point- which refer to the format used to store and manipulate numeric representations of data. The fixed-point DSPs are designed to represent and manipulate a real data type – positive and negative whole numbers – with a fixed number of digits after, and sometimes before, the decimal point via a minimum of 16 bits (2^{16}). On the other hands, the floating-point DSPs represent and manipulate rational numbers via a minimum of 32 bits (2^{32}).
In this paper, a structure of parametric vocal-tract model is proposed, which is formed by combiningIIR and FIR digital filters [2]. The model is optimised with respect to implementation on a fixed-point digital signal processor with Harvard architecture.

APPROXIMATION OF TRANSCENDENTAL FUNCTIONS BY CONTINUED FRACTION EXPANSION

Transcendental functions are usually used for the approximation of the non-linear functions in the digital signal processing domain. The continued fraction expansion is the mostly used approximation for the transcendental functions [3, 4]:

$$e^{x}=\cfrac{1}{1-\cfrac{x}{1+\cfrac{x}{2-\cfrac{x}{3+\cdots+\cfrac{x}{2-\cfrac{x}{2s-1+\cdots}}}}}} \qquad (1)$$

$$=\frac{1}{1}-\frac{x}{1}+\frac{x}{2}-\frac{x}{3}+\cdots+\frac{x}{2}-\frac{x}{2s-1}\cdots$$

Where S refers to an integer number used for computation other coefficient of the fraction expansion.

$$\ln(x)=\frac{2(x-1)}{x+1}-\frac{(x-1)^2}{3(x+1)}-\cdots-\frac{s^2(x-1)^2}{(2s+1)(x+1)}-\cdots \qquad (2)$$

$$\arctan(x)=\frac{x}{1}+\frac{x^2}{3}+\frac{4x^2}{5}+\cdots+\frac{s^2x^2}{(2s+1)}+\cdots \qquad (3)$$

$$\sqrt{x}=1+\frac{x-1}{2}+\frac{x-1}{2}+\frac{x-1}{2}+\cdots \qquad (4)$$

The approximation accuracy of the transcendental function depends on the number of members of the continued fraction expansion.In other words, it depends on the order of *s*. For example, a sequence of rational fractional functions can be written using the approximation of exponential function (1), whose number of members successively increases.In this case, the approximation accuracy increases with a higher *s*:

$$e^x = \frac{1}{1}, \quad \frac{1}{1-x}, \quad \frac{2+x}{2-x}, \quad \frac{6+2x}{6-4x+x^2}, \quad \frac{12+6x+x^2}{12-6x+x^2}, \ldots \tag{5}$$

The functions are also referred to as the Padé approximation of the exponential function. It is recommended to use only an odd order of the approximation in order to have the same order of both the numerator and the denominator of rational fractional function (5).

Considering the set of approximation functions (5), we can suggest the transfer functions as follows:

$$H_1(x) = \frac{2+x}{2-x}, \quad H_2(x) = \frac{12+6x+x^2}{12-6x+x^2}, \quad H_3(x) = \frac{120+60x+12x^2+x^3}{120-60x+12x^2-x^3}, \quad H_4(x) = \frac{1680+840x+180x^2+20x^3+x^4}{1680-840x+180x^2-20x^3+x^4}, \ldots \tag{6}$$

where $x = a\, z^{-1}$, and z is the parameter of the *Z*-transform.

Cepstral Models of Vocal Tract With Both Formants And Anti-Formants

Speech as an analog sound signal is generated by exciting the human vocal tract, which begins in the larynx and ends in the lips. The speech signal can be approximately divided into voiced and un-voiced segments. This signal is time-variant.Thus it is usually segmented into small segments of 10 ms to 20 ms in length, in which speech can be considered as approximately stationary. The magnitude frequency response of the vocal tract shows sections with peaks-resonances (known as formants) and sections without any resonances-valleys (known as anti-formants), in which the response can decrease to zero[11].

The vocal tract models distinguish from each other by approximating only formants (transfer function poles) or both formants and anti-formants (transfer function poles and zeros). Therefore the vocal tract models can be divided into two groups:

The classical Linear Prediction (LP) approachhas been investigated in the literature, in which only formants are approximated[1, 10]. The vocal tract is modeled by an all-pole IIR digital filter with the transfer function [1]:

$$G(z) = \frac{\sqrt{a}}{A(z)}, \tag{7}$$

where $A(z)$ is a polynomial of the order M and $\sqrt{a}$ is the Root Mean Square (RMS) value of the residual (excited) signal. The zeros of $A(z)$ define the poles of the transfer function and correspond to the formants of the vocal tract. To realize a transfer function $G(z)$ a lattice structure with two or four multipliers per one section is used. The lattice structure is robust enough, so it is not necessary to initial the inner state-space variables of the structure used between segments.This is because the partial transient responses caused by single poles or groups of poles are short and they do not cause any problems.

The cepstral model approximates both formants and anti-formants of the vocal tract [2]. The transfer function of the IIR digital filter is :

$$H(z) = b\, \frac{P(z)}{Q(z)}. \tag{8}$$

The polynomial $P(z)$ defines the zeros of the transfer function of the digital filter (anti-formants), and the polynomial $Q(z)$ determines the poles of the transfer function (formants). The constant b defines the input signal volume, in which the vocal tract model is excited. By using the robust structures of the type IIR digital filter we can model the transfer function (8) simply but sufficiently accurately.

Let us consider the logarithmic spectrum $\ln\left|S\left(\mathrm{e}^{\mathrm{j}wT}\right)\right|$ of a short segment of speech signal $\{s[n]\}$, obtained by multiplying a sampled speech signal by a windowing function (most frequently the Hamming window). $T$is the sampling interval and it holds $T=1/f_{\mathrm{s}}$, where $f_{\mathrm{s}}$is the sampling frequency. The angular frequency is defined as $w=2\pi f$. This logarithmicspectrum can be expressed with the aid of real cepstrum $\{c[n]\}$:

$$\ln\left|S\left(\mathrm{e}^{\mathrm{j}wT}\right)\right|=\sum_{n=-\infty}^{\infty}c[n]\mathrm{e}^{-\mathrm{j}nwT}=c[0]+2\sum_{n=1}^{\infty}c[n]\cos nwT\ . \qquad (9)$$

A minimum-phase digital filter, in which the logarithm of the transfer function approximates the envelope of function (9), is defined as follows[2]:

$$\tilde{S}(z)=\mathrm{e}^{c[0]}\mathrm{e}^{2\sum_{n=1}^{N_0-1}c[n]\,z^{-n}}=b\ \mathrm{e}^{2C(z)}, \qquad (10)$$

where $0<N_0<L_{\min}=f_{\mathrm{s}}/f_{0\max}<N_{\mathrm{F}}/2$. The value of $f_{0\max}$ is the maximum pitch frequency of the speech signal, and N_{F} is the number of points obtained from the FFT algorithm. To obtain the original spectrum of the speech signal using the synthesis of cepstralcoefficients $\{c[n]\}$, an exponential function must be realized that is inversed to the logarithmic function. By using the exponential function, the transfer function (10) cannot be realized in real time and it has to be approximated. The multiplication coefficient $b=\mathrm{e}^{c[0]}$ is equal to the mean value of the logarithmic speech spectrum magnitude.

The function*C(z)*is the *Z*-transform of the windowed causal part of real cepstrum $\{c[n]\}$, which in the given segment describes the vocal tract properties, and which corresponds to an FIR digital filter with non-linear phase.

$$C(z)=\sum_{n=1}^{N_0-1}c[n]z^{-n} \qquad (11)$$

The block diagram of the realization of the transfer function (10) is presented in Fig.1.

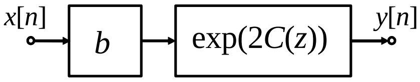


Fig.1. Cepstral vocal-tract model defined by Eq. (10).

This model is not suitable for implementation on a digital signal processor.The rational approximation of the exponential function must be used that would enable an effective realization of the vocal tract model.

CHOICE OF REALIZATION STRUCTURE SUITABLE FOR DIGITAL SIGNAL PROCESSOR IMPLEMENTATION

The exponential function $\exp(C(z))$ can be expressed by means of continued fraction expansion [3]:

$$\mathrm{e}^{2C(z)}=\frac{a_1}{b_1}+\frac{a_2}{b_2}+\frac{a_3}{b_3}+\cdots+\frac{a_s}{b_s}+\cdots\ , \qquad (12)$$

where $a_1=b_1=1$, $a_2=-2C(z)$, $b_2=1+C(z)$, and for $s\geq 3$ $a_s=C^2(z)/(2s-5)(2s-3)$, $b_s=1$.

The continued fraction expansion (12) can be approximated by a set of rational functions called Padé approximants. The s^{th}-order Padé approximant equals:

$$\tilde{G}(z)=\frac{1+a_{s1}C(z)+a_{s2}C^2(z)+\cdots+a_{ss}C^s(z)}{1-a_{s1}C(z)+a_{s2}C^2(z)-\cdots+(-1)^s a_{ss}C^s(z)} \qquad (13)$$

The coefficients a_{si} as well as the stability conditions for $C(z)$ consisting of only one element are summarized in Table 1 .

Table 1. Coefficients of the Padé approximants according to (13)

a_{si}	**1**	**2**	**3**	**4**	**5**	**Stability Conditions**
1	1					$\lvert c(n)\rvert<1$
2	1	1/3				$\lvert c(n)\rvert<\sqrt{3}$
3	1	2/5	1/15			$\lvert c(n)\rvert<2.3$
4	1	3/7	2/21	1/105		$\lvert c(n)\rvert<3$
5	1	4/9	1/9	1/63	1/945	$\lvert c(n)\rvert<3.6$

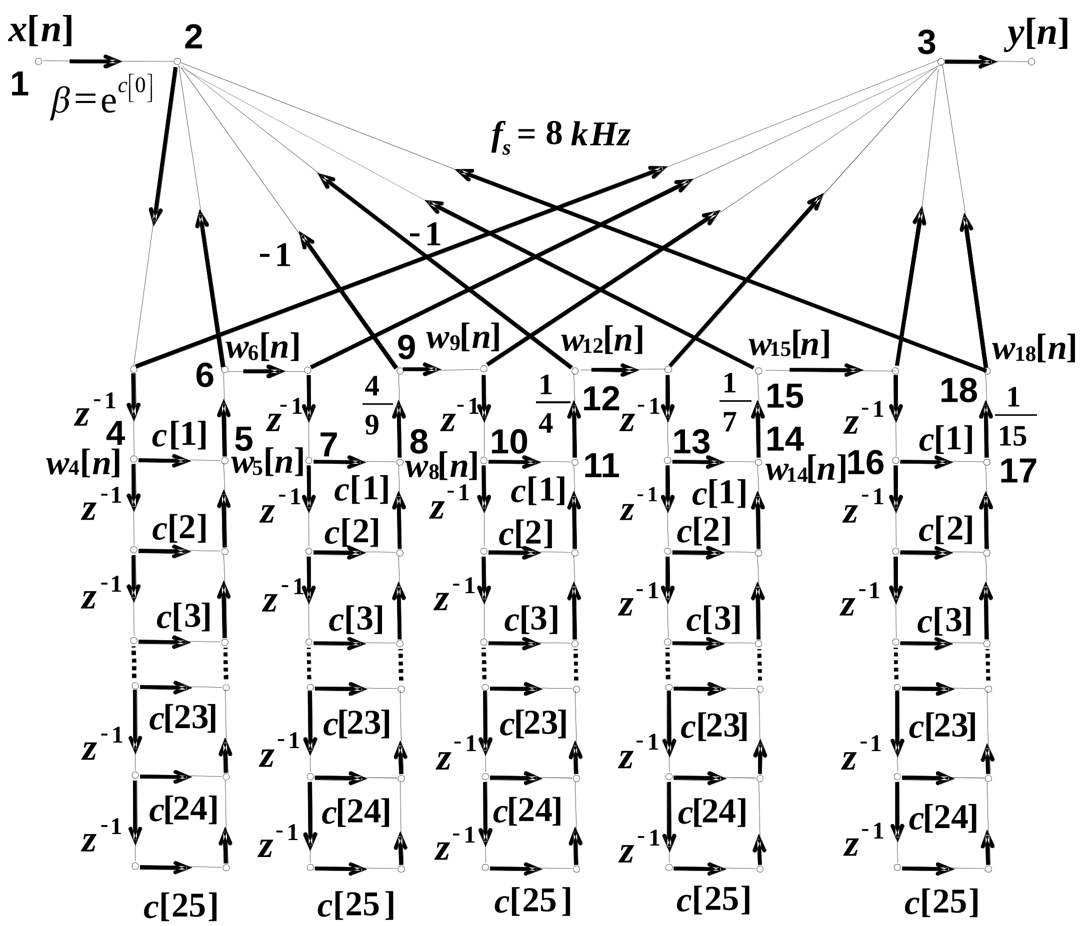


Fig. 2. Cepstral vocal-tract model of the 5th order with N_0 = 26cepstral coefficients (f_s = 8 kHz), realized by an IIR digital filter in 2ndcanonic form. FIR digital filters are also realized in 2ndcanonic form

The transfer function of an already realizable s^{th}-order IIR digital filter, which models function (10) in the form of (13), is presented in (14) [2]:

$$\tilde{\tilde{S}}(z) = b\frac{1+a_{s1}C(z)+a_{s2}C^2(z)+\cdots+a_{ss}C^s(z)}{1-a_{s1}C(z)+a_{s2}C^2(z)-\cdots+(-1)^s a_{ss}C^s(z)} \quad (14)$$
$$=\frac{Y(z)}{X(z)}=H(z) \quad ,b=e^{c[0]},$$

where $Y(z)$ is the Z-transform of output signal y[n], and $X(z)$ is the Z-transform of input signal x[n].

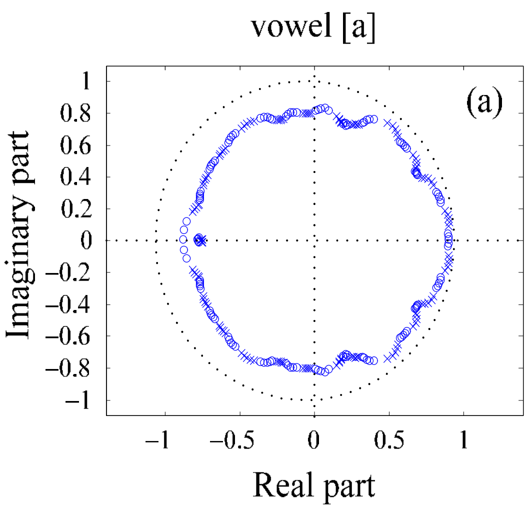
The transfer function (14) can be modified such that state-space difference equations are obtained for implementing the cepstral model on a DSP. The modification in question will concern a type IIR digital filter with transfer function (14) that will contain instead of individual delay blocks z^{-1} the transfer function of an FIR digital filter with non-linear phase defined by equation (11).

Experiments have shown that employing an IIR cepstral vocal tract model of maximally 5th order guarantees sufficient approximation accuracy for both the sampling frequency f_s = 8 kHz (N_0 = 26 cepstral coefficients) and the frequency f_s =16 kHz (N_0 = 52 cepstral coefficients) [2]. Fig. 2 presents the signal-flow graph of the vocal tract model of the 5th order realized by a digital filter of the IIR type, which is implemented in the 2nd canonic form. The reason behind using the 2nd canonic form can be made clear only when the state-space difference equations are being derived.

VERIFICATION OF CEPSTRAL MODEL ACCURACY BY COMPUTER ANALYSIS

To verify the cepstral model accuracy the Matlab script of the semi-symbolic analysis has been programmed [7, 8].

Figs. 3 and 4 show the calculated pole (formants) and zero (anti-formants) plots of transfer function (14) for 26 cepstral coefficients, which have been calculated on the basis of an analysis of the top parts of sounds [a] and [s]. The results respond to assumptions; that is the structure in Fig.2 can be used for an implementation of the vocal tract on a digital signal processor with the Harvard architecture.



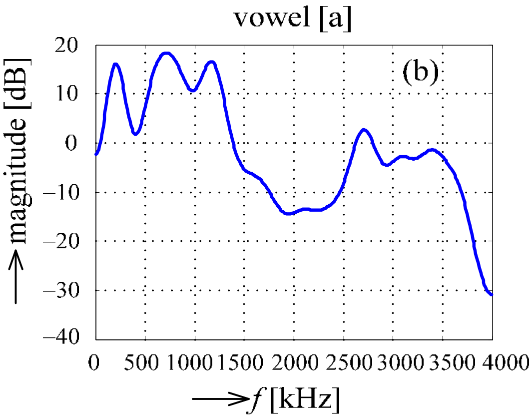


Fig. 3. Zero-pole plot of the transfer function (a) and the magnitude frequency response (b) of the cepstral model of the 125th order vocal tract from Fig.2 for the vowel [a].

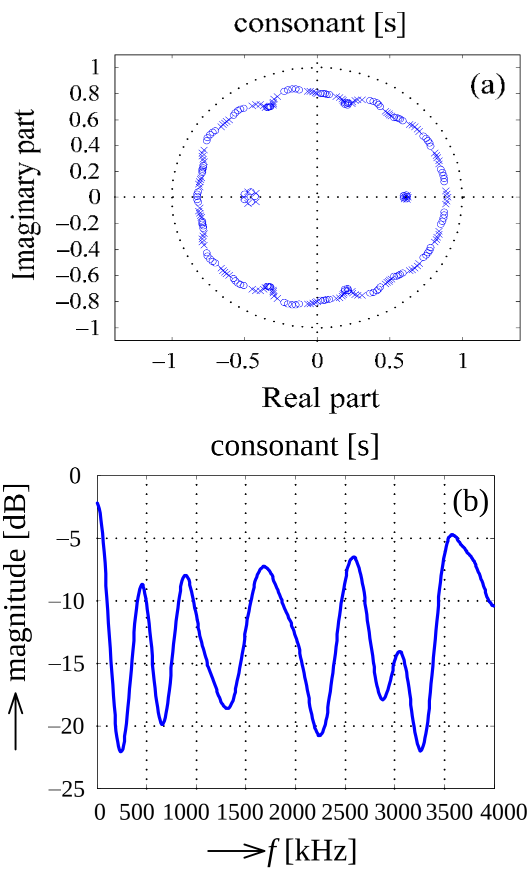


Fig. 4. Zero-pole plot of the transfer function (a) and the magnitude frequency response (b) of the cepstral model of the 125th order vocal tract from Fig.2 for the consosnant [s].

IMPLEMENTATION OF ALGORITHM ONDIGITAL SIGNAL PROCESSOR WITH HARVARD ARCHITECTURE

Considering the second canonical structure of the IIR digital filter, the set of state-space canonical equations starts from transfer function (14) and for the 5th order of the transfer function.

$$\begin{aligned}
&V_1(z)=C(z)W_1(z), \quad W_2(z)=a_{51}V_1(z),\\
&V_2(z)=C(z)W_2(z), \quad W_3(z)=\frac{a_{52}}{a_{51}}V_2(z),\\
&V_3(z)=C(z)W_3(z), \quad W_4(z)=\frac{a_{53}}{a_{52}}V_3(z),\\
&V_4(z)=C(z)W_4(z), \quad W_5(z)=\frac{a_{54}}{a_{53}}V_4(z), \qquad (15)\\
&V_5(z)=C(z)W_5(z), \quad W_6(z)=\frac{a_{55}}{a_{54}}V_5(z),\\
&W_1(z)=b\ X(z)+W_2(z)-W_3(z)+W_4(z)\\
&\qquad -W_5(z)+W_6(z),\\
&Y(z)=W_1(z)+W_2(z)+W_3(z)+W_4(z)\\
&\qquad +W_5(z)+W_6(z).
\end{aligned}$$

The state-space difference equations, which we get from Eq. (15) by using the inverse Z-transform, are as follows:

$$\begin{aligned}
&v_1[n]=c[n]*w_1[n], \quad w_2[n]=a_{51}v_1[n],\\
&v_2[n]=c[n]*w_2[n], \quad w_3[n]=\frac{a_{52}}{a_{51}}v_2[n],\\
&v_3[n]=c[n]w_3[n], \quad w_4[n]=\frac{a_{53}}{a_{52}}v_3[n],\\
&v_4[n]=c[n]*w_4[n], \quad w_5[n]=\frac{a_{54}}{a_{53}}v_4[n], \qquad (16)\\
&v_5[n]=c[n]*w_6[n], \quad w_6[n]=\frac{a_{55}}{a_{54}}v_5[n],\\
&w_1[n]=b\ x[n]+w_2[n]-w_3[n]+w_4[n]-w_5[n]\\
&\qquad +w_6[n],\\
&y[n]=w_1[n]+w_2[n]+w_3[n]+w_4[n]+w_5[n]\\
&\qquad +w_6[n].
\end{aligned}$$

The partial operations between the state-space difference equations express a realization of the FIR digital filter:

$V_i(z)=C(z)W_i(z)$, where$i$ = 1,2,3, ... , s (in our case s = 5) (17)

Since the time-domain convolution corresponds to this multiplication:

$$\begin{aligned}
v_i[n]&=c[n]w_i[n]=\sum_{m=1}^{N_0-1}c[m]w_i[n-m]\\
&=c[1]\,w_i[n-1]+c[2]\,w_i[n-2]+c[3]\,w_i[n-3]+\mathbf{L} \qquad (18)\\
&\quad +c[N_0-1]\,w_i[n-N_0+1]; \quad i=1,2,....,5
\end{aligned}$$

The convolution is also realized in the 2nd canonical structure. The location of the state-space variables and model coefficients in the data memories of Motorola DSP56300 digital signal processor is presented in Fig. 5. The calculation of two loops is described in Fig. 6. The main operation of the inner loop is a multiplication and accumulation instruction. Before starting this loop, accumulator A is set to zero. Two input registers X0 and Y0 are loaded with state-space variable $w_1[n–1]$ (address pointer r0 shows the place where the value is stored) and cepstral coefficient $c[1]$ addressed by pointer r4, respectively. After the values from memories X: and Y: have been read, the address pointers are automatically incremented. The initialized modulo mode guarantees that the values are only within the limits given in Fig. 5. The values in the input registers are multiplied and added to the values in accumulator A. At the same time, other values $w_1[n–2]$ and $c[2]$ are loaded into input registers, and so on. This instruction is applied (N_0–2) times in this loop. After completing the last multiplication and accumulation operations in the inner loop, the last input values $w_1[n–N_0+1]$ and $c[N_0–1]$ are in the input registers. Then these values have to be multiplied and accumulated outside the inner loop. Accumulator A now contains state-space variable$v_1[n]$. This value is first multiplied by two and then by the Padé coefficient rate. In the end the obtained value $w_2[n]$ is loaded into memory X: at the address r0. After the finalization of all operations inside the inner loop, pointer r0 refers to the address of w2[n–1]. Pointer r4 refers to the address of c[1], and pointer r5 represents the address of coefficient

−4/9. The outer loop controls the inner loop and repeats it as many times as is the order of the IIR digital filter (in our case 5 times). After the outer loop is finished, all state-space variables from w2[n] to w6[n] are calculated, which are necessary to obtain both value w1[n] and output signal y[n]. These values are stored in data memory X: and their addresses are N0 positions from each other. Address pointer r0 indicates the value w1[n–1]. Then it is decremented to the position w1[n], which is not known yet. But the addresses in r0 are automatically increased by N0, which is stored in register n0. The recursion of these values that corresponds to the last two state-space equations (16) can now be realized.

X:

$w_5[n-N_0+1]$	$\leftarrow$ r0+5N_0−1
⋮	
$w_4[n-2]$	$\leftarrow$ r0+4N_0+2
$w_4[n-1]$	$\leftarrow$ r0+4N_0+1
$w_4[n]$	$\leftarrow$ r0+4N_0
⋮	
$w_2[n-N_0+1]$	$\leftarrow$ r0+2N_0−1
⋮	
$w_2[n-2]$	$\leftarrow$ r0+ N_0+2
$w_2[n-1]$	$\leftarrow$ r0+ N_0+1
$w_2[n]$	$\leftarrow$ r0+ N_0
$w_1[n-N_0+1]$	$\leftarrow$ r0+N_0−1
$w_1[n-N_0+2]$	$\leftarrow$ r0+N_0−2
⋮	
$w_1[n-2]$	$\leftarrow$ r0+2
$w_1[n-1]$	$\leftarrow$ r0+1

Y:

−1/15	$\leftarrow$ r5+4
⋮	
−4/9	$\leftarrow$ r5+1
−1	$\leftarrow$ r5
$c[N_0-1]/2$	$\leftarrow$ r4+N_0−1
⋮	
$c[3]/2$	$\leftarrow$ r4+2
$c[2]/2$	$\leftarrow$ r4+1

Fig. 5. Lay-out of the state-space variables, the Padé and cepstral coefficients in data memories of the digital signal processor for the 2nd canonical structure of IIR and FIR digital filters.

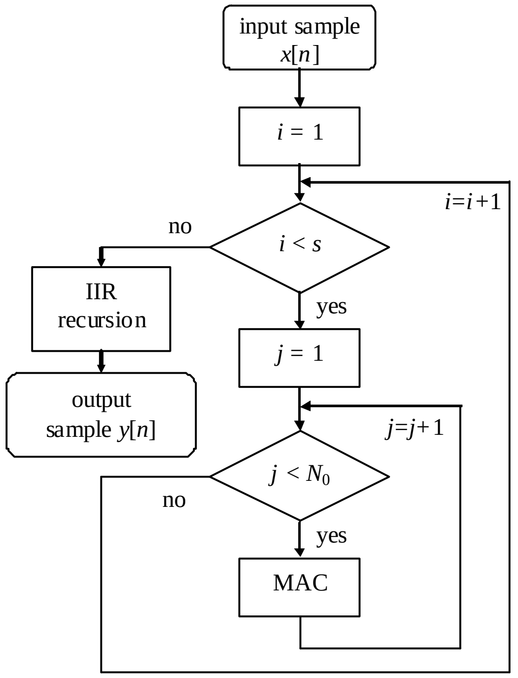


Fig. 6. Block diagram showing the cepstral model calculation in assembler of digital signal processor DSP56300 (Motorola).

The assembly program for digital signal processor DSP56300 of the cepstral vocal tract model is (both IIR and FIR digital filters are realized in the 2nd canonical structure) [9]:

VERIFICATION OF PROPER SYNTHESIS ON DIGITAL SIGNAL PROCESSOR

Figure 7 shows a comparison between the original time signal of vowel [a] and the time signal synthesized by the cepstral model. The Root Mean Square (RMS) error is defined as a geometrical average of the difference between the original magnitude and the magnitude that is being modeled:

$$\mathrm{RMS} = 20\log\left(\sqrt{\frac{\sum_{k=1}^{N_F}\left(\left|H[k]\right|-\left|\tilde{H}[k]\right|\right)^2}{N_F}}\right).\ [\mathrm{dB}] \qquad (19)$$

Spectrum magnitude $H[k]$ at frequency points f_k was determined from an original recording of the vowel [a] by the inverse cepstral transform of the same cepstral coefficients that were used in the model, as presented in Fig. 8. The second spectrum magnitude $\tilde{H}[k]$ was calculated by the same procedure as in the first case, but the synthesized signal was used. An error comparison for various stationary parts of speech sounds is shown in Table 2. The magnitude of the errors is mostly below the 3 dB level. The higher error of the consonant [s] is caused by an unvoiced excitation whose power spectrum density in the scope of the analyzed segment is not absolutely flat. As a result, we can say that the simulation gives good results.

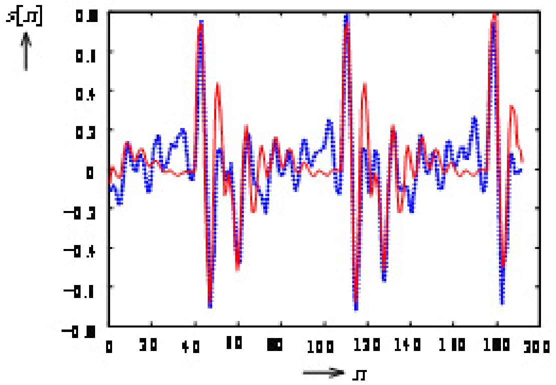


Fig. 7. Comparison of the original time signal of vowel [a] (dashed line) and the synthesised time signal (solid line).

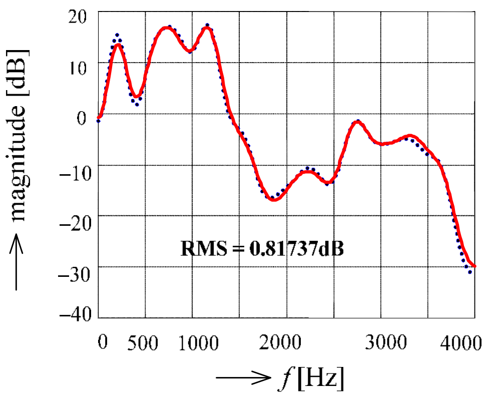


Fig. 8.Magnitude frequency response of the original vocal tract (dashed line), and the magnitude frequency response of the cepstral model (solid line) for vowel [a].

Table 2. RMS errors for various sounds

Vowels and Consonants	Root Mean Square Error (RMSE)
[a]	0.82 dB
[e]	2.58 dB
[i]	1.24 dB
[o]	2.21 dB
[u]	2.31 dB
[m]	3.14 dB
[l]	2.82 dB
[s]	5.13 dB

CONCLUSIONS

A new approach to the transcendental function approximation by the continued fraction expansion is described in the paper. One of the approximations, namely the approximation of the exponential function, is used for vocal tract modelling. The cepstral model is designed to have both formants and anti-formants. The transfer function of the vocal tract is realized as a combination of the FIR and IIR digital filters and the corresponding algorithm is implemented on a Motorola DSP56300 digital signal processor. The algorithm can be modified for implementation on other digital signal processors. Table 2 shows good resultsand there is good agreement between the theoretical and synthesised time signals of chosen speech sounds. According to our results and the results presented in the literature, the designed cepstral model is useful for TTS systems and it gives better results than the LP model, which models only vocal tract formants.

ACKNOWLEDGEMENT

The paper has been supported within the framework of the SIX research project (CZ.1.05/2.1.00/03.0072).

References:

1. VÍCH,R., SMÉKAL,Z. All-Pole and Zero-Pole Speech Modelling (Invited Paper). In Proceedings of the International Conference „BIOSIGNÁL '98". June 23-25, 1998, Brno, Czech Republic, pp.196-199. ISBN 80-214-1169-4
2. VÍCH,R. PŘIBIL,J. SMÉKAL,Z,: New Cepstral Zero-Pole Vocal Tract Models for TTS. In Proceedings of the International Conference EUROCON '2001, July 7-9, 2001, Bratislava, Slovakia, pp.459-462.
3. KHOWANSKYI, A.N.: Application of Continued Fractions and Their Generalizations in Numerical Analysis. State Publishing House for Engineering and Theoretical Literature. Moscow 1956. (In Russian)
4. VÍCH, R., SMÉKAL, Z.: Digital Filter Realization of Non-rational Transfer Functions. In Proceedings of the EURASIP Conference "Signal Analysis and Prediction I", Prague, June 24-27, 1997, pp. 179-182.
5. VÍCH,R., SMÉKAL,Z. Digital Filters. Academia, Prague 2000. (In Czech) ISBN 80-200-0761-X
6. SMÉKAL Z., VÍCH R.: Cepstral Speech Synthesis Optimised for Dual Harvard Architecture of DSP. In: Proceedings of the International Conference on Telecommunications (ICT' 2000)", May, 2000, Acapulco, pp.244-248. ISBN 968-36-7762-2
7. TASSIGNON, H., SMÉKAL, Z. Semi-Symbolic CAD Discrete System AnalysisInProceedings of the 16th International Conference on Electronic Speech Signal Processingjoined with 15th Czech-German Workshop "Speech Processing".Dresden: TUD Press, Verlag der WissenschaftenGmBH, 2005, pp. 466-471.
8. SMÉKAL, Z. AL-KHEIR, J.: Semi-symbolic CAD Discrete System Analysis and Signal Flow Graphs. In Proceedings of the 30th International Conference on Telecommunications and Signal Processing TSP 2007. September 2007. BUT Brno, Czech Republic, pp. 93-96. ISBN 378-8021-43445-5
9. SMÉKAL, Z., SYSEL, P. Digital Signal Processors. Prague: Sdelovaci technika, 2006. (In Czech) ISBN: 80-866-4508-8
10. SMÉKAL, Z., VONDRA, M., VÍCH, R., State-Space Representation of CepstralVocal Tract Model for DSP Implementation Internet Journal ELECTRONICSLETTERS, 3/9/2002.ISSN 1213-161X
11. John R. Deller, Jr., John H. L. Hansen, John G. Proakis: Discrete time processing of speech signal, J.-Willy-Intercience 1993, ISBN 0-7803-5386-2

Receieved 1/4/2012