

Research on Dependences of Speech Pitch Parameters on Pulse and Heartbeat Signals

Dmitry Poleshenkov
ITMO University
49 Kronverksky Pr., St. Petersburg, 197101, Russia
d.poleshenkov@yandex.ru

Ekaterina Pakulova
Southern Federal University
105/42 Bolshaya Sadovaya Str., Rostov-on-Don, 344006, Russia
epakulova@sfedu.ru

Oleg Basov
ITMO University
49 Kronverksky Pr., St. Petersburg, 197101, Russia
oobasov@corp.ifmo.ru

Abstract

In this paper, we consider the influence of the cardiovascular system to the process of speech production. We propose the mathematical model of the impact of cardiovascular system elements on the process of speech pitch synthesis. This model takes into account the effect of the functioning of the heart and the pulsations of the large vessels of the thorax on the intensity of the air flow from the lungs during breathing, as well as the effect of the pulsations of the blood vessels of the vocal cords on the speech synthesis process. We made the initial check of approximate properties of the model in practice. In the future, the proposed model allows to construct the precision model of speech signal forming. The solution of the direct synthesis problem allows to increase the speech quality in the synthesis, recognition and coding processes. The inverse problem aims to increase the efficiency of physiologic and psycho-emotional state estimation of a person. In conclusion, we discussed further research directions that may improve the quality of the estimates obtained on the basis of the formulated model.

1 Introduction

Analysis of mathematical models of the processes of speech production [FF14, LS17, LZ16, RN19, Sor85, Sor16] in the tasks of recognition of a physiological and psycho-emotional state of a user of a socio-cyber-physical system

Copyright 2019 for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

In: S. Hölldobler, A. Malikov (eds.): Proceedings of the YSIP-3 Workshop, Stavropol and Arkhyz, Russian Federation, 17-09-2019–20-09-2019, published at <http://ceur-ws.org>

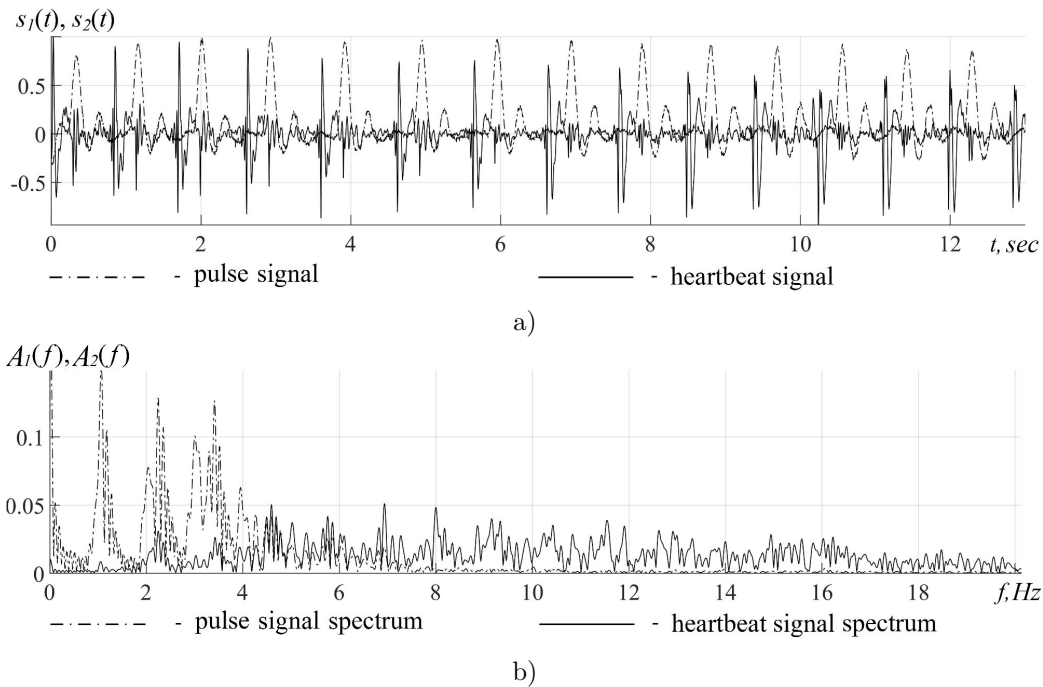


Figure 1: Signals of pulse and heartbeat (a) temporal representation; b) spectral representation)

shows that they are very specific. The consideration of functioning of vocal tract uncoupled from other human body systems leads to significant deviation of modelling results from the real physical process.

One of such human body systems that has a significant influence on speech production is the cardiovascular system. In [Bab05, BO14] is shown that speech pitch period contains periodic and random components. In [BO14] authors also suppose that periodic changes of pitch may be caused by blood flow pulsation.

The contribution of the present work is detection of dependency between changes of pitch parameters and functioning of the cardiovascular system in the process of speech production

2 Analysis of human pulse and heartbeat

The cardiovascular system influences on all body systems. Its work is mainly characterized by heart and distal vessels. Periodic atrial and ventricular contraction of a heart with vessels guarantees blood motion.

A normal heart rate for adults ranges from 65 to 85 beats per minute. The cardiac cycle consists of two periods: the period of atrial and ventricular contraction (systole) and the period of their relax (diastole). Duration of systole is on average 0,33 sec, duration of diastole is 0,47 sec [PV03].

The temporal representation of the pulse (pulsogram) $s_1(t)$ (see Fig. 1a) has two peaks in the period, corresponding to the moments of release of blood from the ventricles and the reflection of blood flow from the closed semilunar heart valves, while the pause between these peaks is fixed. Pulse spectral bandwidth $A_1(f)$ (see Fig. 1b) on average doesn't exceed 20 Hz [PA12].

Cardiophonography is one of the methods for the analysis of heart functioning which records the heart sounds. The temporal representation of heartbeat (phonocardiogram) $s_2(t)$ is different from pulsogram (see Fig. 1a). It caused by the complex process of heart functioning. The signal is periodic with distinct peaks corresponding to the first (systole) and second (diastole) tones of heart rate. The maxima of these tones fall at the closure of the atrioventricular and semilunar valves, respectively. In the spectral representation of the heartbeat signal $A_2(f)$ (see Fig. 1b), an increase in the amplitudes of the spectral components is observed in the band from 4 to 20 Hz with respect to the spectral representation of the pulse signal.

The phonocardiograms are mainly characterized by the sounds of the heart valves and the blood flow in heart chambers, the aorta and the pulmonary artery. In this regard, the mechanical vibrations transmitted to the surface of the lungs will differ significantly from the corresponding acoustic signal of the heartbeat. It is proved by a change in the structure of recorded phonocardiograms depending on the signal recording point [Tsa18].

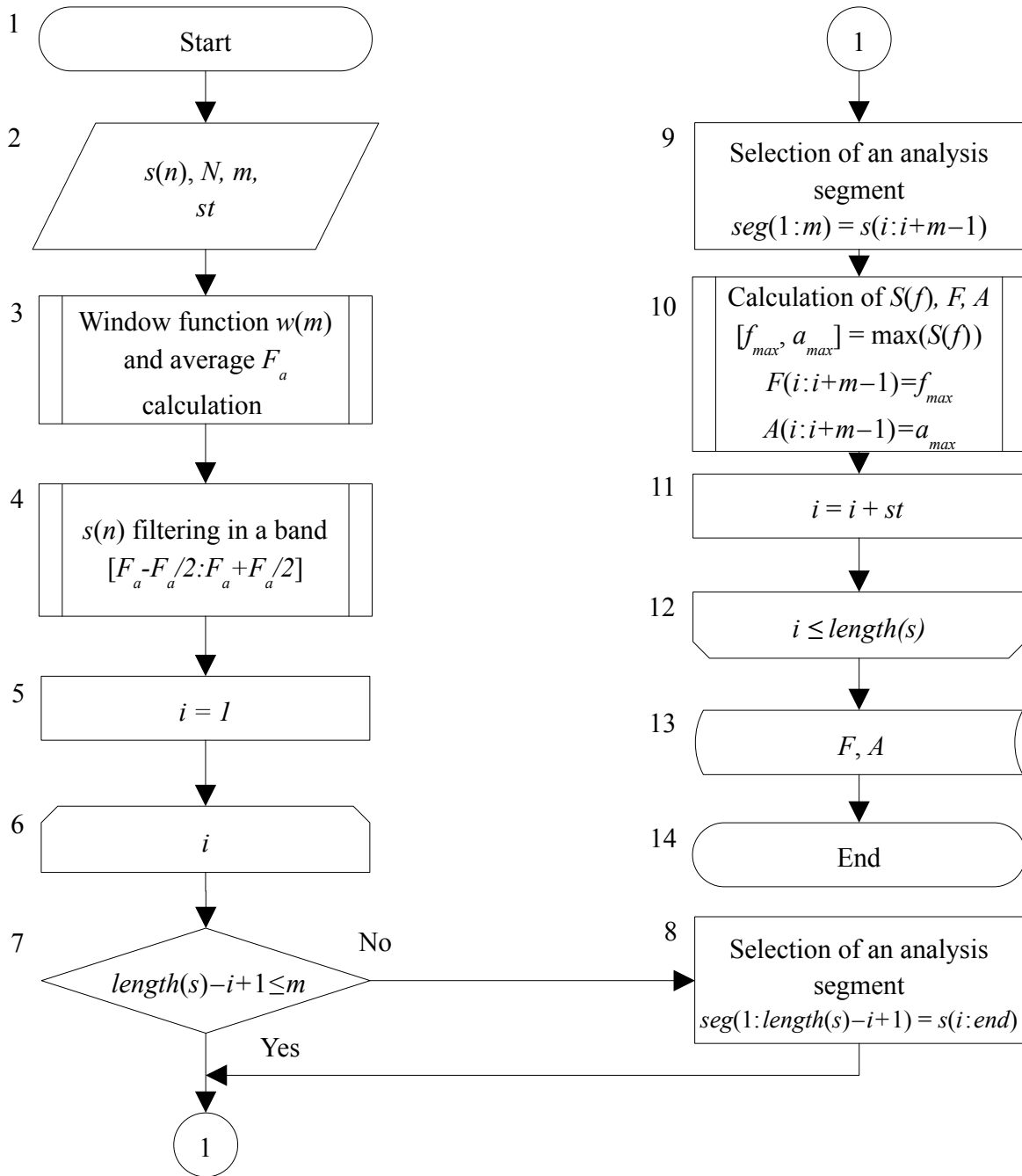


Figure 2: Algorithm of the pitch trajectories separation

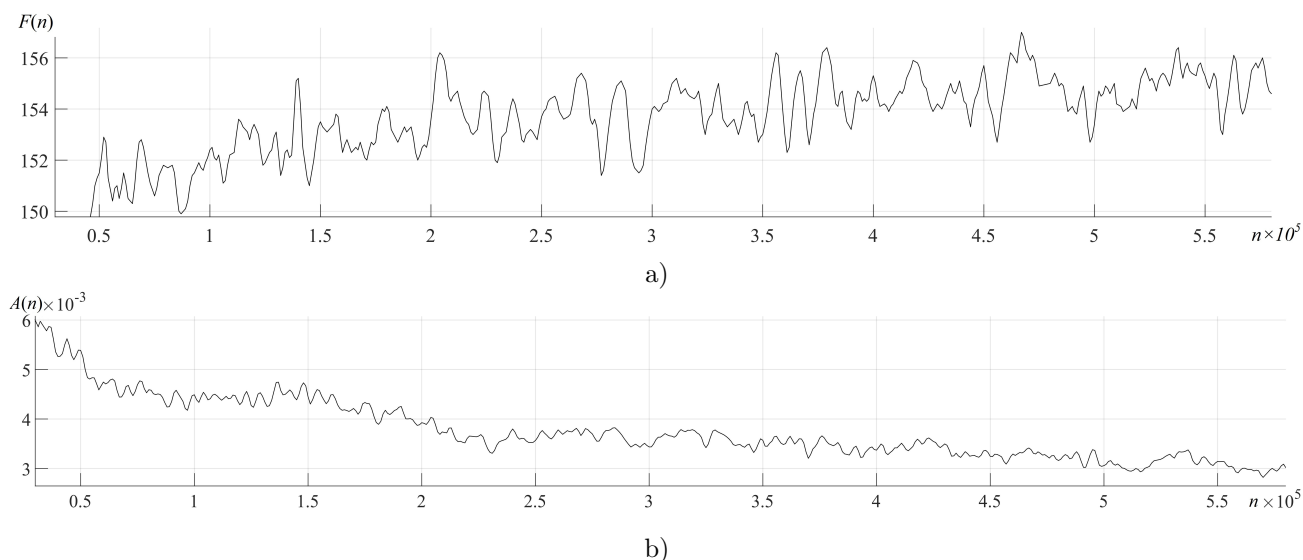


Figure 3: The results of the pitch trajectories separation algorithm (a - the pitch frequency trajectory; b - the pitch amplitude trajectory)

3 Description of mechanisms of cardiovascular system influence on the synthesis process of speech pitch

Based on the physiology of the blood circulation process and the structure of the organs of the circulatory system [PM85, Kab19], we may assume the following basic mechanisms of the influence of the circulatory system on the synthesis of the pitch of speech:

- the effect of heart contractions on the intensity of the air flow from the lungs by physical impact on their surface;
- the effect of changes in blood flow in the aorta, pulmonary artery and blood vessels of the lungs on the intensity of the air flow;
- the effect of pulsations of the blood vessels of the vocal cords on the phase relationships of the acoustic wave process.

Since the heart is located close to the surface of the lungs, heart contractions have a periodic mechanical impact on them. It leads to the intensity fluctuation of the air stream during the process of breathing. By the physical model of speech production [Sor92], a change in the flow intensity leads to a change in the frequency and amplitude of the pitch during speech synthesis. It is worth also to note that both atrial contraction and ventricular contraction will cause fluctuations in the intensity of the air stream.

Similarly, changes in airflow during breathing are affected by pulsations of the aorta, pulmonary artery, and pulmonary blood vessels. Given the close relative positioning of the organs of the cardiovascular system inside the thoracic cavity, the total mechanical effect on the lungs is complex.

During the assessing the influence of the pulsation of the small blood vessels of the vocal cords on the process of speech production, one should take into account their small size and low blood pressure in them [Sud00, VS17]. From this, we can conclude that the pulsations of these blood vessels cause a periodic change in the oscillation phase of the vocal cords. Herewith we should take into account the delay of pressure fluctuations in the vessels of the vocal cords in relation to the heart contractions.

4 Analysis of frequency and amplitude variations of pitch

In order to get the most accurate conclusion about the effect of the cardiovascular system on the change of pitch parameters, it is necessary to minimize the distortion of the analysed speech signals in the pitch frequency band (50–350 Hz). To do this we study sound records of long-spoken vocalized phonemes (which were synchronously recorded with pulsograms and phonocardiograms) with the following parameters: sampling frequency $f_d = 48$

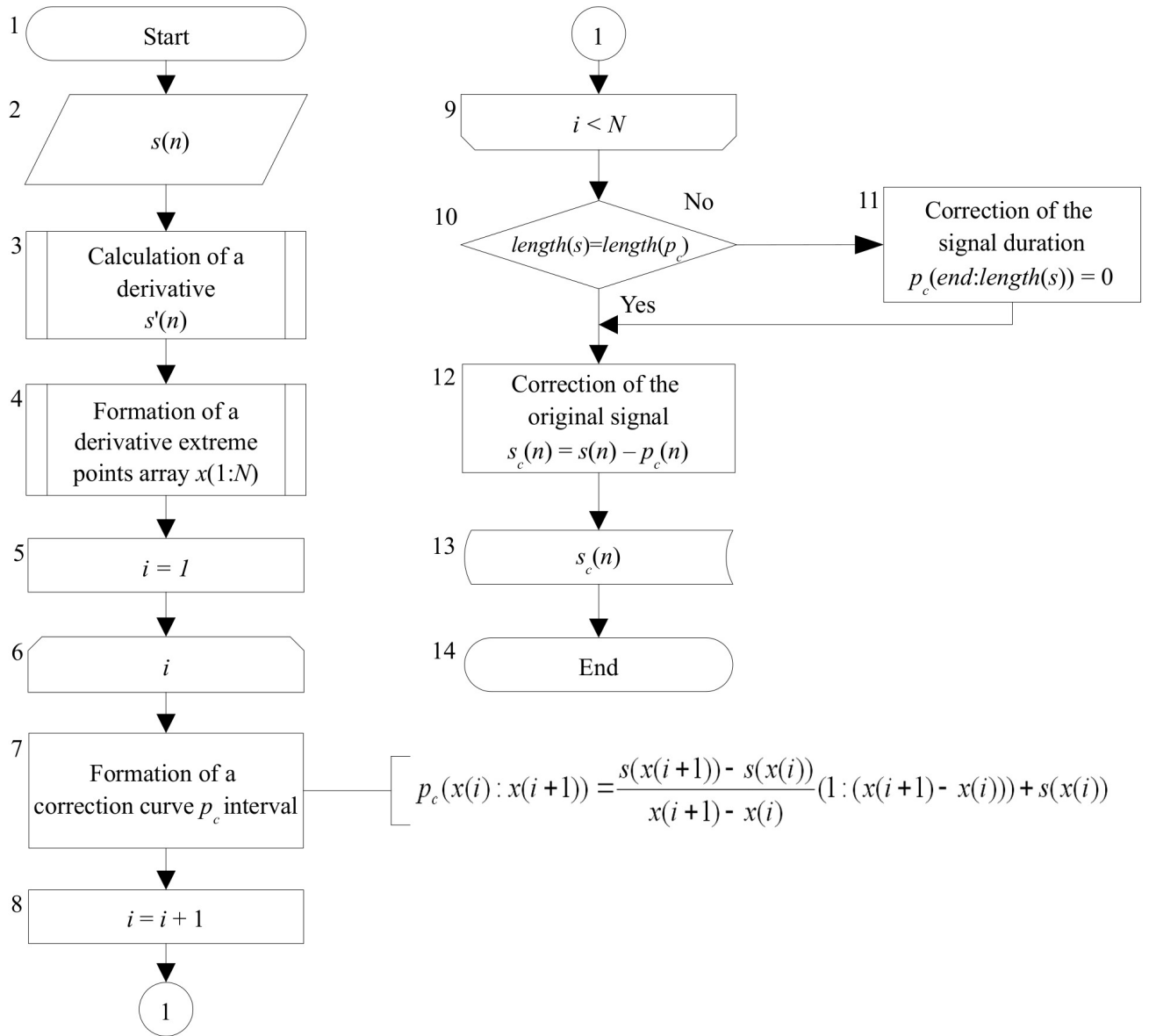


Figure 4: Correction algorithm of the pitch trajectories

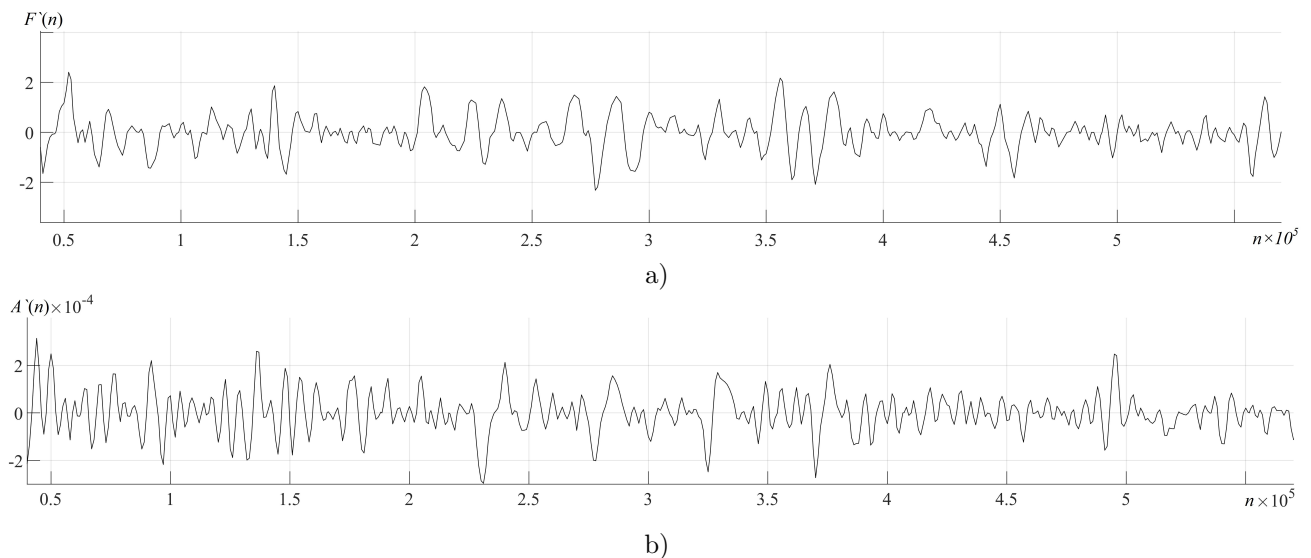


Figure 5: The results of the correction algorithm (a - the trajectories of the pitch frequency; b - the trajectories of the pitch amplitude)

kHz; bit rate $B = 1,536$ bps; microphone bandwidth $2\Delta f = 20$ kHz; duration $t = 10 \dots 14$ s. The use of such initial signal allows us to take the influence of the articulation component beyond the scope of the study, which, makes it possible to abandon the complex ways of recording the speech signal.

During the analysis of common approaches to the selection of pitch parameters [DHKM18, PPS18, SSGZ16, VM16, ZL18], the algorithm based on the spectral separation method (see Fig. 2) was selected. This algorithm allows to obtain an estimate of the pitch frequency and amplitude variations with sufficient time resolution determined by the dimension of the discrete transform Fourier, transform window width and signal sampling rate. The principle of the algorithm is based on the samples allocation of the frequency trajectories and the amplitude of the pitch calculated at intervals of analysis frames based on the spectral representation $S(f)$. The proposed algorithm is simplified and does not imply the extraction of phonemes from the speech signal since there are no phoneme transitions in the studied signals.

The algorithm uses the following input data: $s(n)$ is the analysed speech signal; N is the dimension of the Fourier transform; m is the duration of the analysis time window; st is the step shift of the analysis window. In order to increase the accuracy of trajectory extraction, the shift step is equal to 1. The shift step of the analysis window should be less than the length of the pitch period. This eliminates the necessity to synchronize the analysis window with the pitch period and removes the need to use electroglotograms. The output data of the algorithm are samples sequences of the frequency trajectories $F(n)$ (see Fig. 3a) and amplitudes $A(n)$ (see Fig. 3b) of the pitch. During our research of vocalized phonemes pronounced by different speakers, the result of the algorithm (in terms of pitch frequency trajectories) completely coincides with the result of frequency demodulation of the 1st harmonic of the speech signal.

The received signal of the pitch frequency trajectory carries a large constant component (average pitch frequency), which makes it difficult to assess the dependence of the considered oscillation on the processes of the cardiovascular system. Additionally, there is the presence of low-frequency components caused by the intonation of the spoken segments, as well as a change of the lungs volume as it exhales, which negatively affects the accurate analysis. To compensate for the influence of these factors, we apply the correction algorithm to minimize the distortion (see Fig. 4).

The input data for the algorithm is only corrected signal $s(n)$. The principle of the correction algorithm is based on calculating the intervals of the correction curve p_c on the signal duration between the extreme points $x(1 : N)$ and the derivative $s'(n)$ of the corrected signal, succeeded by subtracting the generated curve from the original signal. The algorithm introduces additional distortions caused by the uneven (piecewise linear) nature of the correction curve. However, they lead to an insignificant spreading of the periodic components of the spectrum of the signal, since the position of the zero crossing points and the position of the signal extremes are not changed relative to the time scale.

For the analysis of the samples sequence of the trajectories of the frequency $F'(n)$ (see Fig. 5a) and the

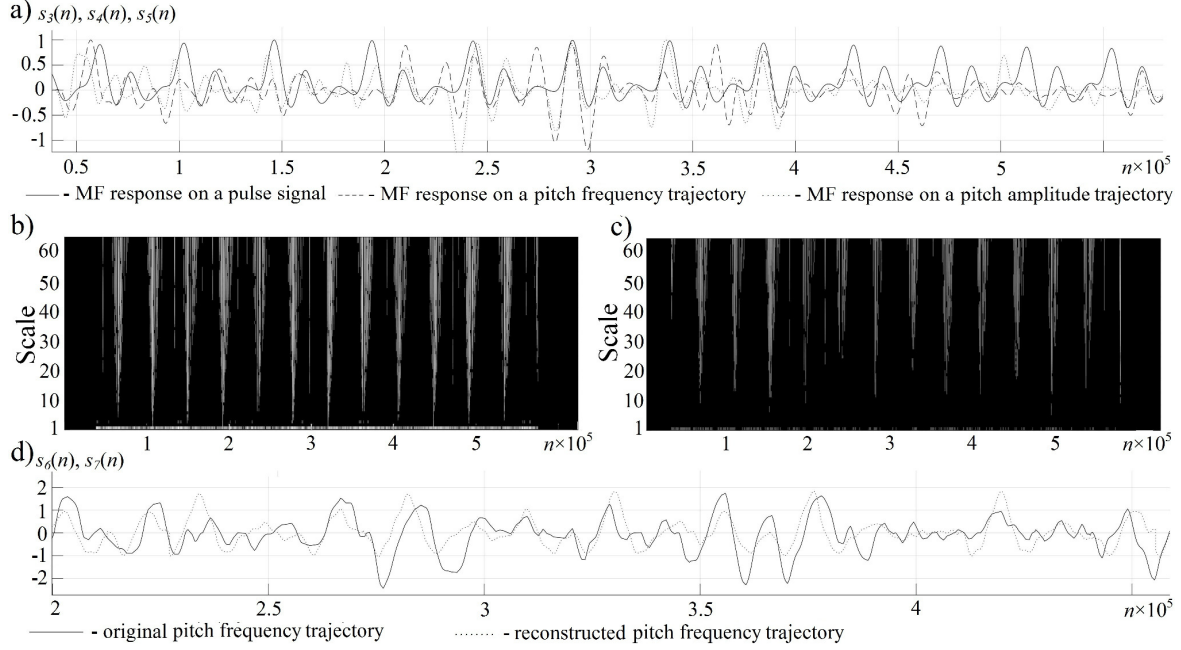


Figure 6: Results (a) responses of the MF, b) wavelet transform coefficients for pitch frequency trajectory, c) wavelet transform coefficients for pitch amplitude trajectory, d) results of signal reconstruction

amplitude $A'(n)$ (see Fig. 5b) of the pitch for the presence of components caused by the cardiovascular system, we made an assessment using a matched filtering mechanism. The reflected periods of the pulse signal were used as the impulse response of the filters. We also analyse these signals with the wavelet transform.

5 Results

As a result of matched filtering, we obtain the periodic signals $s_4(n)$ (response of the matched filter (MF) to the signal of the pitch frequency trajectory), $s_5(n)$ (response of the MF to the signal of the pitch amplitude trajectory (see Fig. 6a)). In general, the position of the corresponding local extremes of these signals (with a small delay) corresponds to the responses $s_3(n)$ of the MF. These responses are getting by filtering the signals of the temporal representation of the pulse.

Since the total mechanical effect transmitted to the surface of the lungs is complex and differs from the recorded pulsograms and phonocardiograms, the responses of the MF to the signals of the frequency trajectories and amplitude of the pitch have a complex structure.

Additional experiments aimed at research on the changes in the pressure of the air stream during breathing will allow to evaluate the total impact. We used wavelet transform (Daubechies wavelet) for the analysis that showed the presence of periodic changes in the trajectories of the frequency and amplitude of pitch, corresponding to heart contractions (see Fig. 6b, Fig. 6c).

Based on the studied mechanisms of the influence of the functioning of the cardiovascular system on the process of pitch synthesis and the obtained experimental data, the pitch frequency trajectory for vocalized phonemes can be determined as follows.

$$f(t) = F_0 + f_0(t) + m_1 \frac{dp(t - \tau_1)}{dt} + m_2 h(t - \tau_2) \quad (1)$$

where $f(t)$ is the pitch frequency trajectory, F_0 is the average frequency of the pitch, $f_0(t)$ is the intonation component, m_1 , m_2 are proportionality factors, τ_1 , τ_2 are the time delays, $p(t)$ is the effect on the vocal cords, $h(t)$ is the total effect on the lungs from the organs of the cardiovascular system located inside the thoracic cavity.

In accordance with expression 1 (see Fig. 6d), the reconstructed signal $s_6(n)$ repeats the extremes of the pitch frequency trajectory $s_7(n)$. A model of aperiodic oscillations of the resonant system with a large attenuation coefficient was used as a model of impact on the lungs from the organs of the cardiovascular system located

Table 1: Summarized results of correlation analysis

Sex	Sample size	Low bound of correlation coefficient B_{min}	High bound of correlation coefficient B_{max}	Average correlation coefficient B_{av}
Male	50	0.67	0.8	0.72
Female	50	0.68	0.77	0.75

inside the thoracic cavity. Moments of oscillations correspond to ventricular and atrial systoles. An unmodified pulse signal was used as a signal to influence the vocal cords.

In order to assess the repeatability of the presented data, we check them with the speech signal (male and female) of long-spoken vocalized phonemes. We estimate the correlation coefficient corrected in accordance with the correction algorithm (see Fig.5) and the signal restored by the expression (1). The summarized results are shown in table 1.

6 Conclusion

The results presented in this paper suggest that the cardiovascular system makes the main contribution to the change of the frequency and amplitude of the pitch period. It allows us to construct the precision model of a speech signal forming that takes into account the influence of the considered factors. The solution of the direct synthesis problem allows to increase the speech quality in the synthesis, recognition and coding processes. The inverse problem aims to increase the efficiency of the physiologic and psycho-emotional state estimation of a person.

7 Acknowledgements

The authors would like to thank the anonymous referees for their valuable comments and helpful suggestions. This work is supported by the Russian Foundation For Basic Research (grant №18-07-00380A).

References

- [Bab05] V.V. Babkin. Noise immune speech pitch isolator. In *Proceedings of the 7th international conference "Digital signal processing and its application"*, volume X-1. IPU RAS, 2005.
- [BO14] Shalaginov V.A. Basov O.O., Nosov M.V. Pitch-jitter analysis of the speech signal. In *SPIIRAS Proceedings*, volume 32, 2014. on russian.
- [DHKM18] Thomas Drugman, Goeric Huybrechts, Viacheslav Klimkov, and Alexis Moinet. Traditional machine learning for pitch detection. *IEEE Signal Processing Letters*, 25(11):1745–1749, 2018.
- [FF14] Mohamed Hesham Farouk and Farouk. *Application of wavelets in speech processing*. Springer, 2014.
- [Kab19] N.A. Kabanov. *Human Anatomy*. Urait Publishing House, 2019.
- [LS17] AS Leonov and VN Sorokin. Upper bound of errors in solving the inverse problem of identifying a voice source. *Acoustical Physics*, 63(5):570–582, 2017.
- [LZ16] NA Lyubimov and EV Zakharov. Mathematical model of acoustic speech production with mobile walls of the vocal tract. *Acoustical Physics*, 62(2):225–234, 2016. on russian.
- [PA12] Ompokov V.D. Pavlov AE, Boronoev V.V. The research of the level of fitness of sportsman organism at the diagnostic complex apdk. *Journal of Buryat State University*, 4, 2012.
- [PM85] Bushkovich V.I. Prives M.G., Lisenkov N.K. *Human Anatomy*. Izdatelstvo Meditsina, 1985.
- [PPS18] Monisankha Pal, Dipjyoti Paul, and Goutam Saha. Synthetic speech detection using fundamental frequency variation and spectral features. *Computer Speech & Language*, 48:31–50, 2018.
- [PV03] Korotko G.F. Pokrovsky V.M. *Human physiology*. Izdatelstvo Meditsina, 2003.

- [RN19] K Sreenivasa Rao and NP Narendra. *Source Modeling Techniques for Quality Enhancement in Statistical Parametric Speech Synthesis*. Springer, 2019.
- [Sor85] V.N. Sorokin. *Theory of speech production*. Radio i sviaz, 1985. on russian.
- [Sor92] V.N. Sorokin. *Speech synthesis*. Izdatelstvo Nauka, 1992. on russian.
- [Sor16] VN Sorokin. Segmentation of the period of the fundamental tone of a voice source. *Acoustical Physics*, 62(2):244–254, 2016. on russian.
- [SSGZ16] Michael Staudacher, Viktor Steixner, Andreas Griessner, and Clemens Zierhofer. Fast fundamental frequency determination via adaptive autocorrelation. *EURASIP Journal on Audio, Speech, and Music Processing*, 2016(1):17, 2016.
- [Sud00] K.V. Sudakov. *Physiology. Basics and functional systems*. Izdatelstvo Meditsina, 2000. on russian.
- [Tsa18] V.P. Tsarev. *Auscultation of the heart*. Belarussian state medical university, 2018. on russian.
- [VM16] DA Volf and RV Meshsheryakov. Model of process of singular estimation of the primary tone of a speech signal. *Acoustical Physics*, 62(2):215–224, 2016.
- [VS17] D.S. Sveshnikov V.M. Smirnov, V.A. Pavdivtseva. *Physiology: A textbook for students of medical and pediatric faculties*. Izdatelstvo MIA, 2017. on russian.
- [ZL18] Xiaoheng Zhang and Yongming Li. Pitch tracking algorithm based on evolutionary computing with regularisation in very low snr. *The Journal of Engineering*, 2018(16):1509–1514, 2018.