Cepstrum-Based Pitch Detection of Industrial Product Sound

Dejan Ćirić, Marko Janković, Marko Milenković, Miljan Miletić

Abstract—**Various audio features can be extracted from audio signals. One of very important ones is pitch. Different algorithms and methods have been proposed in literature to detect pitch. Among them, cepstrum-based pitch detection as a frequency domain method has often been used in practice. Cepstrum is calculated as the inverse Fourier transform of the logarithm of signal spectrum. The fundamental frequency and pitch in this way is estimated as the maximum value of cesptrum in the defined segment. Here, pitch of some industrial products (compressors and DC motors) are estimated by applying the modified cepstrum-based algorithm. The detected pitch values can be used to make a distinction between different working conditions of these products such as different rotation-per-minute (rpm).**

Index Terms—**Cepstrum analysis; Pitch detection; Peak finding; Audio Feature; Audio signals; Industrial product sound.**

I. Introduction

PITCH detection is a common task present in a number of researches mostly related to speech, since the pitch (or fundamental frequency) is one of the most important parameters of speech. Thus, detection of pitch can be found in different speaker recognition and identification systems, speech synthesis systems, telecommunication systems, etc. [1-3]. In addition, pitch is one of the audio features (attributes) used for audio classification, detection and recognition by applying machine or deep learning [4].

Pitch can be detected in the time or frequency domain. One of simple time domain algorithms (methods) is the zero crossing rate method. The most important methods in the time domain are typically based on auto-correlation using a hypothesis that the auto-correlation function of a periodic signal is also periodic and that these two periods are coincident [5].

Regarding the frequency domain methods for pitch detection, one of the most popular is cepstrum-based method. Power cepstrum has some similar properties with the complex cepstrum obtained by homomorphic deconvolution [5], with the main difference that phase information is lost in the power cepstrum, which is called in the rest of the paper cepstrum.

Cepstrum, complex cepstrum, and homomorphic deconvolution have been applied in various areas such as audio processing, speech processing, geophysics, radar, medical imaging, etc. [5]. Some of the applications of both cepstrum and complex cepstrum include restoration of old phonograph recordings [6], cepstral pitch detector, speech recognition and speaker identification.

The periodicity present in a signal that is related to the pitch can be estimated from the cepstrum. Comparing with some other methods for pitch detection, the cepstrum method is able to provide accurate and robust results, but it is computationally complex [5].

This paper presents potentials for using pitch as an audio feature of sound of certain industrial products such as compressors and DC motors. The goal is to investigate if such a feature is able to provide clear distinction between different compressors or DC motors. The pitch is estimated by using cepstrum-based algorithm modified in a sense that peak finder is applied to the obtained cepstrum. Different compressors are related to compressors working with different rotations-per-minute (rpm), while different DC motors are related to different types of DC motors used in the automotive industry.

II. Pitch Detection Algorithms

Pitch is an important attribute of some audio signals such as speech signals. In speech, as a consequence of the vocal fold vibrations, the signal waveform contains certain periodicity translated into “pitch step” in the time domain and “fundamental frequency” or pitch in the frequency domain. However, pitch as an audio feature can also be of significance in machine and deep learning applied to a variety of audio signals including those containing sounds of industrial products, e.g. DC motors, home appliances or internal combustion engines of passenger vehicles [7].

There are various algorithms for pitch detection divided according to different criteria. Thus, there are block based and event based algorithms [9]. In the block based algorithms, the signal is sliced into small segments assuming that the pitch remains constant during the segments. On the other hand, event based algorithms use pitch marking or epoch detection. Here, pitch is not assumed to be constant over several pitch cycles. This is why these algorithms are able to track fast pitch changes even during the segments [10].

According to domain in which the algorithms are applied,

Dejan Ćirić is with the University of Niš, Faculty of Electronic Engineering in Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia (e-mail: dejan.ciric@elfak.ni.ac.rs).

Marko Janković is with the University of Niš, Faculty of Electronic Engineering in Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia (e-mail: marko.p.jankovic@elfak.rs).

Marko Milenković is with the Faculty of Arts, University of Niš, Kneginje Ljubice 10, 18105 Niš, Serbia (e-mail: lionheartmarko@gmail.com).

Miljan Miletić is with the Technical College of Applied Studies in Kragujevac, section in Kruševac, Kosančićeva 36, 37000 Kruševac, Serbia (e-mail: miljan.miletic@vhts.edu.rs).

they can be divided in three groups: time domain, frequency domain and hybrid group (combining time and frequency domain) [11]. Algorithms in the time domain use characteristics of a signal in the time domain, e.g. amplitude fluctuations, zero-crossing and auto-correlation attributes. This group contains the algorithms such as parallel processing time-domain method [12], data reduction method [13], modified auto-correlation (AUTOC), average magnitude difference function (AMDF) and YIN [14].

Frequency domain algorithms are based on the property that periodicity in the time domain is translated into series of peaks (impulses) in the frequency domain. This group of algorithms contains methods such as harmonic product spectrum (HPS), cepstrum-based pitch detection, linear predicting coding (LPC) and sawtooth waveform inspired pitch estimator (SWIPE) [15].

Hybrid detectors combine both the time and frequency domain algorithms. In that regard, frequency domain algorithms can yield temporary spectral aligned sound waves, and after that, auto-correlation methods are applied to determine the pitch period [11]. Hybrid pitch detection methods include pitch estimation filter with amplitude compression (PEFAC) [16], YAAPT [17], multi-band summary correlogram (MBSC) [18] and BaNa [19].

A. AUTOC as Time Domain Method

Among the time domain pitch detection methods, the most used one is the autocorrelation approach. It is based on finding the highest value of the auto-correlation function. Here, the auto-correlation function (AF) of a signal $s(n)$ (n=0, 1, … , N-1) is defined as

$$AF(k)=\frac{1}{N}\sum_{n=0}^{N-1-k}s(n)s(n+k),\quad k=0,1,...,N-1, \tag{1}$$

where N is the signal (or frame) length, while k is the lag index. The pitch is detected at the location of the peak of auto-correlation function.

B. Cepstrum-based Algorithm as Frequency Domain Method

Cepstrum $C(m)$ can be calculated as the inverse Fourier transform of the logarithm of Fourier transform of the target signal, $s(n)$:

$$\begin{aligned} S(k)&=\log\left\{\sum_{n=0}^{N-1}s(n)\cdot\exp^{-j\frac{2\pi}{N}nk}\right\} \\ C(m)&=\frac{1}{N}\left\{\sum_{k=0}^{N-1}S(k)\cdot\exp^{j\frac{2\pi}{N}mk}\right\} \end{aligned}. \tag{2}$$

The pitch is detected at the location of cepstrum maximum calculated as given in (2). It is presented in the literature that the cepstrum-based method is sensitive to noise in the target signal [11].

C. PEFAC as Hybrid Method

The pitch in the PEFAC algorithm is detected by convolving the power spectral density of the signal in the log-frequency domain with the filter summing the energy of the pitch harmonics [11]. The model at the time moment t of a perfectly periodic signal (having fundamental frequency f_0) in the power spectral density domain can be expressed as

$$Y_t(f)=\sum_{k=1}^{K}a_{k,t}\delta(f-kf_0)+N_t(f), \tag{3}$$

where $N_k(f)$ is the power spectral density of the undesired noise and $a_{k,t}$ is the power of the k-th harmonic. The signal model gets the following form in the logarithmic domain

$$Y_t(f)=\sum_{k=1}^{K}a_{k,t}\delta(q-\log(k)-\log(f_0))+N_t(f), \tag{4}$$

where $q=\log(f)$. The energies of the signal components in this domain can be combined by convolving $Y_t(q)$ with the impulse response filter.

The filter $h(q)$ can suppress the noise with smoothly varying spectra, but this is not the case for high amplitude narrowband noise. This is why the spectrum compression is applied before the convolution with the filter $h(q)$

$$Y_t'(q)=Y_t(q)^{a_t(q)} \tag{5}$$

where t represents time index and $a_t(q)$ represents the compression exponent [11].

III. Methods Applied

For the purpose of carrying out this research, audio signals with different spectral contents are selected. Some of them have known pitch and harmonics distribution, such as the trumpet sound given in Fig. 1, while in the other signals (containing the sounds of certain industrial products) these parameters are unknown. The sound of trumpet is tonal sound with pronounced pitch, and the main characteristic of such a sound is periodicity, which can be seen in the time domain as presented in Fig. 2.

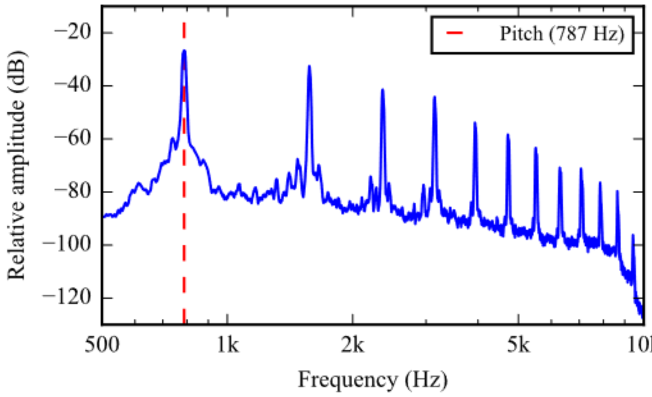

Fig. 1. Spectrum of trumpet sound consisting of fundamental (f_0) at frequency of 787 Hz and partials that have a harmonic distribution relative to the fundamental.

Cepstrum-based pitch detection uses signal periodicity. In this case, the periodicity refers to a particular waveform of the specific length that is repeated throughout the signal, and it is reflected in a discrete spectrum containing prominent peaks equally distributed throughout the frequency range.

The cepstrum-based pitch detection algorithm consists of five main steps described below. In the first step, the signal is transformed from the time to the frequency (spectral) domain using the Fourier transform. By applying the relevant function in Python, two variables are obtained - spectrum (given in complex numbers) and frequency vector. In the second step, the logarithm of the spectrum magnitude is calculated. The basic idea of cepstrum is to transfer the periodicity from the logarithmic representation of the spectrum to the time domain. For that reason, in the third step, the inverse Fourier transform is applied over the logarithm of the spectrum magnitude data. The cepstrum of the trumpet sound from Figs. 1 and 2 obtained in the described way is shown in Fig. 3.

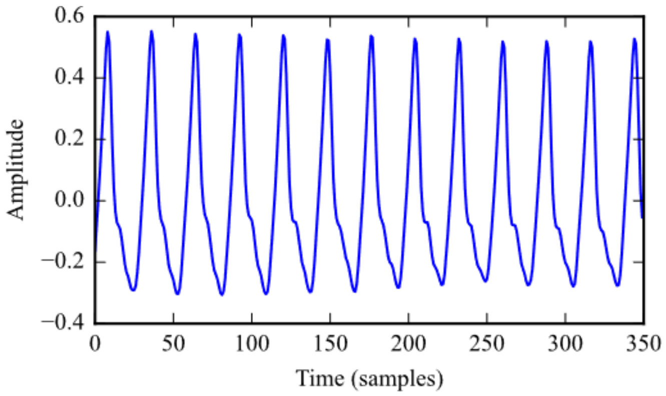

Fig. 2. Trumpet sound shown in time domain.

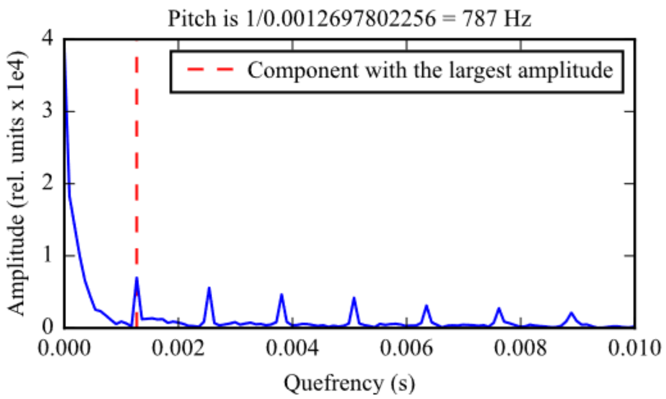

Fig. 3. Cepsrum of trumpet sound.

As can be seen in Fig. 3, the *x*-axis is titled quefrency, which is an anagram of frequency and it is related to time scale. The reason is in performing the inverse Fourier transform, that is, in inverting the frequency by this transformation. As a consequence, the highest frequencies are located at the beginning of the *x*-axis, while the lowest frequencies are located at the end of the *x*-axis. Frequencies are converted into quefrencies by taking 1/frequency.

The fourth step is related to finding the maximum value of the cesptrum, that is, finding the quefrency value at which the cepstrum maximum occurs. In majority of cases, the maximum amplitude of the cepstrum is located at the beginning of the x-axis – at the zero quefrency or its vicinity, which greatly complicates the pitch detection in an automated manner. There are several approaches to overcome this problem. One of them is to limit the quefrency band in which the search of cepstrum maximum is carried out. In this way, an error can be introduced in estimating the pitch of some sounds, such as sound of certain industrial products, since it is not known in advance in which frequency band to expect the pitch. The second approach is not to detect the pitch at the absolute maximum of the cepstrum, but instead to skip the cepstrum maximum at the zero quefrency, and to use the second largest value of the cepstrum for the pitch estimation. Such an approach leads to an error of pitch detection in the case where the cepstrum do not have the maximum value at the zero quefrency.

In this paper, an alternative approach to solve the mentioned problem is applied. Thus, the function for finding peaks (*find_peaks*) from the library *scipy* in *Python* is used. This function finds all local maxima by simple comparison of neighboring values, with the ability to define large number conditions for the peak properties. Due to the fact that there are no neighboring values on the left side of the cepstrum maximum at the zero quefrency (or in close vicinity), this maximum is automatically not considered as a peak.

When the relevant maximum of the cepstrum is selected, the quefrency value of that maximum is converted into frequency representing the estimated pitch (pitch = 1/(quefrency of cepstrum maximum)).

To check the described algorithm, the fundamental frequency from the trumpet sound is removed by filtering, see the spectrum shown in Fig. 4.a). The determined maximum of the cepstrum is located at the same quefrency position, see Fig. 4.b), as before removing the fundamental component.

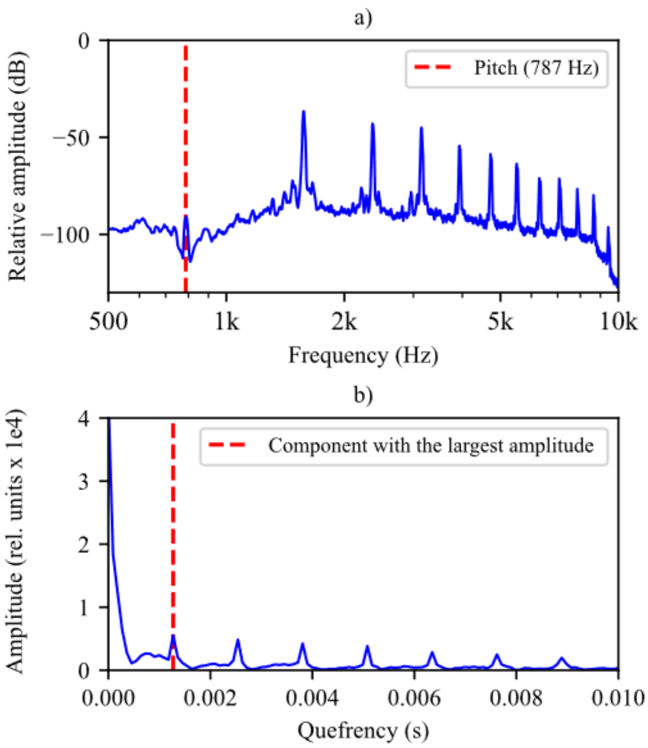

Fig. 4. Spectrum a) and cepstrum b) of trumpet sound after removing the fundamental component located at 787 Hz.

For the purpose of this research, the estimation of pitch by the modified cepstrum-based algorithm described above is done on different sounds of industrial products including fridge compressors and DC motors. The sound of fridge compressors were recorded in the semi-anechoic chamber in three working cycles (modes of operation) having different rpms (4000 rpm, 2400 rpm and 1300 rpm). Sounds of DC motors were also recorded in the semi-anechoic chamber (not the same one used for the compressor recording) within the production line. The recording was done on two different types of DC motors (here denoted type A and type B), in two different directions of rotation (here denoted direction 1 and direction 2) and for two different conditions regarding the failure (without failure and with certain failure). The analysis of recorded audio signals is carried out using the scripts developed in Python 3.8.

IV. Analysis of Detected Pitch

Th

e potentials for applying the cepstrum-based pitch detection in making difference between different samples or working conditions of certain industrial products is investigated here focusing on spectrum and cepstrum of the sounds of these products, that is, on one-figure value of the detected pitch.

The first product whose sound is analyzed is fridge compressors. The target for this product is to consider if it is possible to make a distinction between three different working conditions of compressors – three different rpms (4000, 2400 and 1300). The spectrum and cepstrum of the compressors having 4000 rpm are shown in Fig. 5. The pitch estimated by the described procedure is 6400.05 Hz.

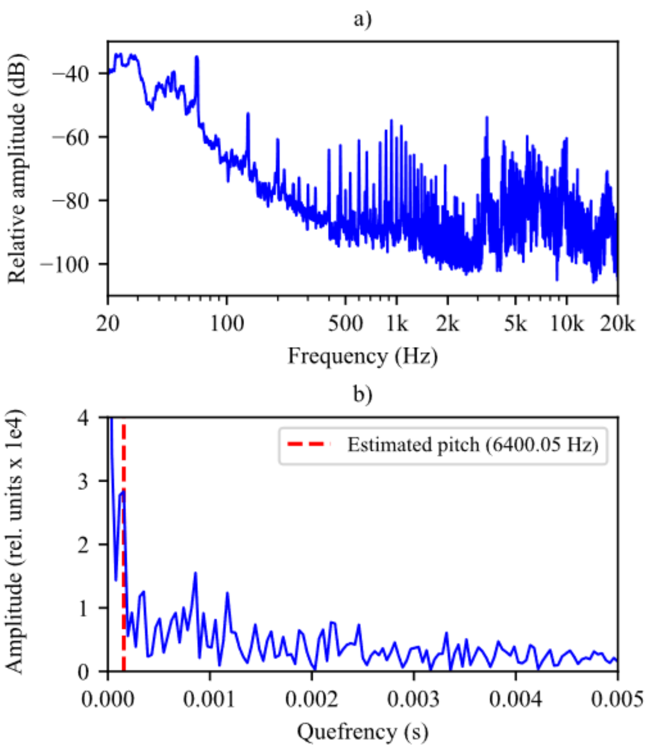

Fig. 5. Spectrum a) and cepstrum b) of fridge compressor sound at 4000 rpm.

When the rpm is changed from 4000 to 2400, it causes certain changes in the spectrum, but also in the cepstrum and consequently in the detected pitch. Fig. 6 shows the spectrum and cepstrum of the fridge compressor sound at 2400 rpm, where the estimated pitch is 2756.25 Hz.

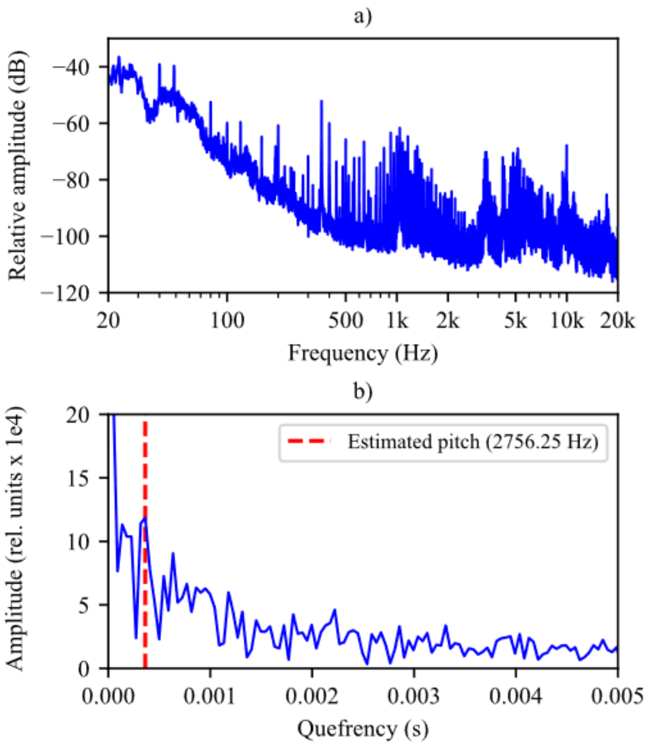

Fig. 6. Spectrum a) and cepstrum b) of fridge compressor sound at 2400 rpm.

By changing the rpm to 1300, the detected pitch is changed to 5120.04 Hz. The spectrum and cepstrum for that rpm are presented in Fig. 7. Regarding the cepstrum, not only the relevant maximum of the cepstrum, but also its pattern is changed by changing rpm. This is why it seems reasonable to introduce at least one more attribute that will reflect dissimilarity of the cepstrum pattern. This attribute can be related either to decay of the cepstrum envelope, cepstrum energy within certain qefrency limits or even a vector containing the cepstrum values within pre-defined qefrency limits.

The next step in the analysis includes DC motors. Spectrum and cepstrum of the same DC motor of type A, but in two opposite directions of rotation are shown in Fig. 8. Comparing the spectra, it can be concluded that overall pattern is similar for both directions of rotation, with certain differences in particular frequency bands and at particular frequencies. The patterns of cepstrum are also similar, but still having some differences for different directions of rotation. The estimated pitch for both directions is the same, 2666.33 Hz, since the maximum value of cepstrum is located at the same quefrency.

Comparison of spectra and “cepstra” of sounds of two DC motors of different types (A and B) is presented in Fig. 9. Shapes of the spectra given in Fig. 9.a) and Fig. 8.a) for the motor A are similar, with certain differences, since different motors of the same type A are used for the analysis. The same situation exists in the "cepstra" from Fig. 8.b) and Fig. 9.b). In spite of these differences, the estimated pitch for two motors of the same type A is the same (2666.73 Hz), while the estimated pitch for the motor of type B is 888.91 Hz.

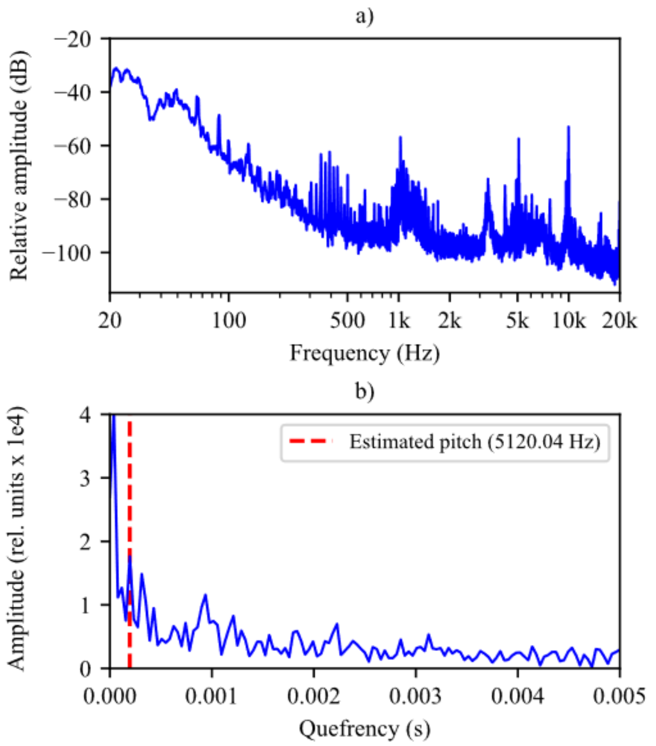

Fig. 7. Spectrum a) and cepstrum b) of fridge compressor sound at 1300 rpm.

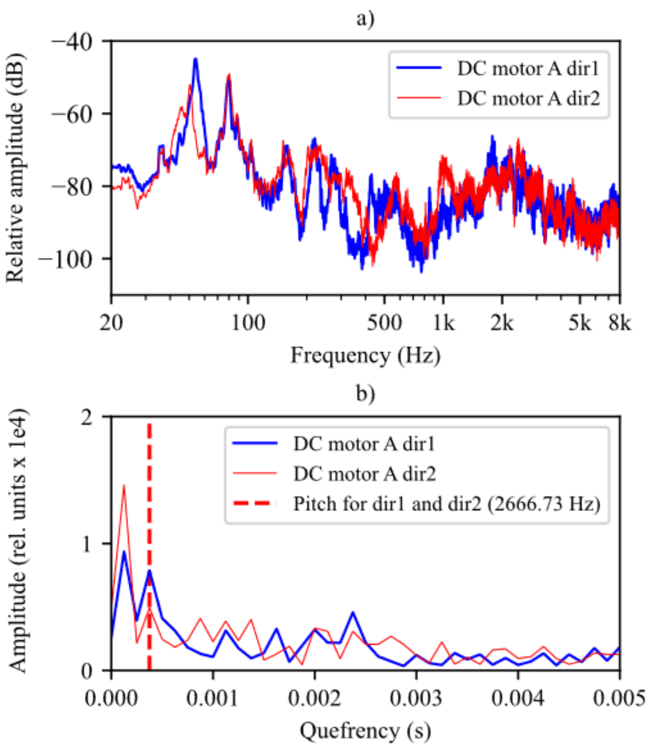

Fig. 8. Spectrum a) and cepstrum b) of DC motor A in two opposite directions of rotation.

When detected pitch values for DC motors with and without failures are analyzed, the differences between motors depend on the failure itself. In some cases, the failure causes change of periodicity or pseudo-periodicity of sound waveform leading to a certain change of the estimated pitch. Fig. 10 illustrates one of such cases presenting spectra and "cepstra" of DC motors without failure (OK motor) and with failure (NOT OK motor). The estimated pitch for OK motor is 2666.73 Hz and for NOT OK motor is 1333.36 Hz. Here, the differences between the spectra and "cepstra" are larger in comparison to the previously analyzed two cases of DC motors.

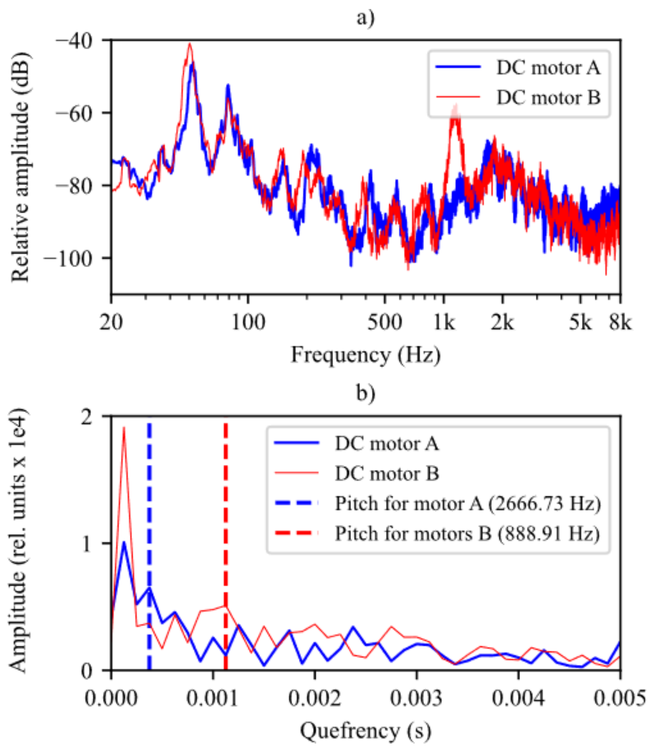

Fig. 9. Spectrum a) and cepstrum b) of different DC motors (motors of type A and B) and direction of rotation 1.

V. Conclusion

This paper analyses potentials for making a difference between working conditions or states of two industrial products, fridge compressors and DC motors, based on pitch of their sounds. The pitch is detected using the cepstrum-based algorithm.

The results show that there are conditions and states where the estimated pitches are different for different conditions (states). However, there are also cases where the pitch only could not be used for making a difference between conditions (states) of these products. Even in such cases, the patterns or shapes of the "cepstra" show certain differences for different conditions (states). This is why a measure calculated from the cepstrum can be introduced in addition to pitch that can be used as a new audio feature.

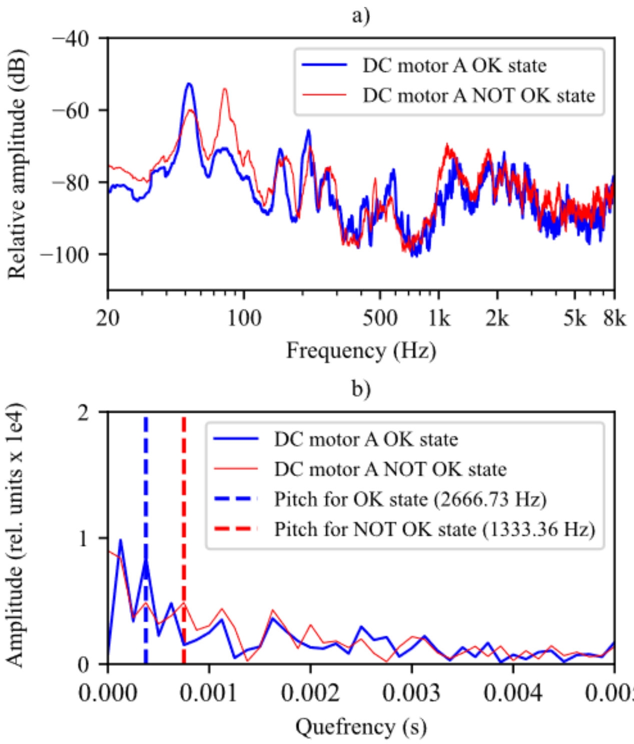

Fig. 10. Spectrum a) and cepstrum b) of OK DC motor (without failure) and NOT OK DC motor (with failure).

Acknowledgment

This research was supported by the Science Fund of the Republic of Serbia, 6527104, AI-Com-in-AI.

References

[1] S. Makov, A. Minaev, I. Grinev, D. Cherhyshov, B. Kudruavcev, V.Mladenovic, "A spectral-based pitch detection method," Proc. AIP Conference (XV International Scientific-Technical Conference "Dynamic of Technical Systems"), Rostov-on-Don, Russia, 2188, 050005, 11-13 September, 2019.

[2] B. Gold, N. Morgan, D. Ellis, *Speech and Audio Signal Processing: Processing and Perception of Speech and Music*, 2nd ed. USA: Willey, 2011.

[3] X. Zhao, Y. Shao, D. Wang, "CASA-Based robust speaker identification," *IEEE Transactions on Audio Speech and Language Processing*, vol. *20*, no. *5*, pp. *1608-1616*, July, 2012.

[4] A. B. Nielsen, L. K. Hansen, U. Kjems, "Pitch based sound classification," Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toulouse, France, vol. 3, 14-19 May, 2006.

[5] K. Pratibha, H.M. Chandrashekar, "Estimation and tracking of pitch for noisy speech signals using EMD based autocorrelation function algorithm," Proc. 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, India, pp. 2071-2075, 19/20 May, 2017.

[6] T.G. Stockham, Jr., T.M. Cannon, R.B. Ingebretsen, "Blind deconvolution through digital signal processing," *Proc. IEEE*, vol. 63, pp. *678-692*, April, 1975.

[7] M. Milivojčević, D. Ćirić, "Extracting relevant acoustic features of the internal combustion engines," Proc. 7th International Conference on Electrical, Electronic and Computing Engineering (IcETRAN), Belgrade, Čačak, Niš, Novi Sad, Serbia, pp. 49-53, 28-29 September, 2020.

[8] A. V. Oppenheim, R. W. Schafer, "From frequency to quefrency: A history of the cepstrum," *IEEE Signal Processing Magazine*, pp. *95-106*, September, 2004.

[9] H. Chander, B. Singh, R. Khanna, "Short time cepstrum analysis method for pitch estimation of an arbitrary speech signal using MATLAB," *International Journal of Research in Electronics and Computer Engineering*, vol. *5*, no. *3*, pp. *1-4*, July-September, 2017.

[10] T. Ananthapadmanabha, B. Yegnanarayana, "Epoch extraction of voiced speech," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. *ASSP-23*, no. *6*, pp. *562-575*, December, 1975.

[11] L Sukhostat, Y. Imamverdiyev, "A comparative analysis of pitch detection methods under the influence of different noise conditions," *Journal of Voice*, vol. *29*, no. *4*, pp. *410-417*, July, 2015.

[12] B. Gold, L. Rabiner, "Parallel processing techniques for estimating pitch periods of speech in the time domain," *Journal of the Acoustical Society of America*, vol. *46*, no. *2*, pp. *442-448*, August, 1969.

[13] N. J. Miller, "Pitch detection by data reduction," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. *23*, no. *1*, pp. *72-79*, February, 1975.

[14] A. de Cheveigne, H. Kawahara, "YIN, a fundamental frequency estimator for speech and music," *Journal of the Acoustical Society of America*, vol. *111*, no. *4*, pp. *1917-1930*, April, 2002.

[15] A. Camacho, "SWIPE: A sawtooth waveform inspired pilch estimator for speech and music," Ph.D. dissertation, University of Florida, Gainesville, Florida, USA, 2007.

[16] S. Gonzalez, M. Brookes, "A pitch estimation filter robust to high levels of noise (PEFAC)," Proc. 19'th European Signal Processing Conference - EUSIPCO, Barcelona, Spain, pp. 451-455, 29 August – 02 September, 2011.

[17] K. Kasi, S. A. Zahorian, "Yet another algorithm for pitch tracking," Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing - ICASSP, Orlando, Florida, USA, 13-17 May, 2002.

[18] L. N. Tan, A. Alwan, "Multi-band summary correlogram-based pitch detection for noisy speech," *Speech Communication*, vol. *55*, pp. *841-856*, 2013.

[19] H. Ba, N. Yang, I. Demirkol, W. Heinzelman, "BaNa: a hybrid approach for noise resilient pitch detection," Proc. IEEE Statistical Signal Processing Workshop (SSP), Ann Arbor, Michigan, USA, 04-07 August, 2012..