A Novel Approach on Speaker Gender Identification and Verification Using DWT First Level Energy and Zero Crossing

Abdelkader Amraoui
Laboratory of Applied Automation and Industrial Diagnostics, Faculty of Sciences and Technology, Ziane Achour University of Djelfa, Djelfa, Algeria
kader2717@yahoo.fr

Slami Saadi
Department of Computer Sciences, Faculty of Exact Sciences and Informatics, Ziane Achour University of Djelfa, Djelfa, Algeria
s.saadi@univ-djelfa.dz

Received: 18 August 2022 | Revised: 3 September 2022 | Accepted: 5 September 2022

Abstract-The aim of this work is to find a new criterion for determining a range of values in order to determine the gender of a speaker. The use of the Discrete Wavelet Transform (DWT) of the Daubechies db7 parent wavelet and the computation of the zero crossing energy from the first level of the DWT was followed by computation of the values of the criterion for both genders and comparison with the value of the speech basic frequency for both genders for the same sign or sentence. The standard has a limited range of values close to the basic frequency range of the same speaker through which we can determine gender. This criterion has been tested on several men and women databases with different repeated sentences for the same person or for both genders and it gives acceptable results that can be worked on.

Keywords-speaker gender; DWT; energy; zero crossing

I. INTRODUCTION

The difference in the linguistic characteristics of humans is characterized by the variance in frequencies of the two genders. Generally, the frequencies of women are smaller than those of men. Usually, the sound consists of vibrations with different lengths and heights. As the human ear cannot hear all the vibrations, some of them are audible, i.e. from 20Hz to 20kHz, and some are inaudible. Frequencies smaller than 20Hz are inaudible and frequencies larger than 20kHz are inaudible and painful to the human ear. Of course, as we get older, the limit of hearing drops, sometimes to 17kHz [1]. The speech signal is non-stationary, complex, and variable with time [2]. In addition, the highest frequency of the speech signal is in the order of 5kHz. The operations done on this signal, such as sampling, require frequencies according to Shannon's law that the sampling frequency is more than twice the frequency of the original signal under study. Many research works have formerly addressed the study and determination of the range of f_0 values [3], which dealt with extracting the basic frequency in the domains of time and frequency as well as wavelets. In addition, some research works deal with a comparison between estimation methods for computing the fundamental for a newborn. In [2], the authors presented techniques for determining the frequency position in time. Other research works [4-6] were interested by this subject, on which researches are still developing since it is not possible to confirm definitively the criterion that determines the gender of the speaker according to the basic frequency domain. Authors in [7] present an efficient approach for automatic speaker identification based on cepstral features, the Normalized Pitch Frequency (NPF) with Discrete Cosine Transform (DCT) and wavelet de-noising pre-processing. On the other hand, wavelet based features in combination with Spectral-Subtraction (SS) were proposed in [8] for speaker identification in clean and noisy environment. Using neural networks for speech recognition tasks, where words are constructed from sequential individual and text-independent sound segments was addressed in [9] for speaker identification. To reduce memory utilization for speaker audio segment identification, authors in [10] proposed a new on line speaker identifier model using short input audio segments. Extracting features from raw speech that captures the unique characteristics of each speaker is accomplished by the filter bank-based Mel Frequency Cepstral Coefficients (MFCC) approach [11] using Discrete Wavelet Transform (DWT). The Average Framing Linear Prediction Coding (AFLPC) technique combined with wavelet transform for text-independent speaker identification systems was presented in [12]. Extracting features from raw speech that capture the unique characteristics of a particular individual using Wavelet Packet Transform (WPT) is presented in [13]. Authors in [14] identify speakers basing on cepstral feature strategy and the NPF as a new feature, for enhancing accuracy using Neural Networks (NN). A motivating application of speech signals identification, for the detection of people with heart failure by glottal features, was presented in [15]. A robust feature extraction method for a real-time speech recognition hardware system was presented in [16].

In the current work, a novel approach for speaker gender identification and verification is introduced. This new criterion is based on DWT and the intersections with zero of the DWT first level for both genders with comparison to speech fundamental frequency. The simulation results prove that the proposed approach enhances the performance of speaker

Corresponding author: Slami Saadi

identification, especially with the DWT and the de-noising pre-processing step.

II. SPEECH SIGNAL CHARACTERISTICS

The speech signal is an audio carrier of complex and diverse information and has features such as the fundamental frequency characteristic, which is often denoted by f_0. Other characteristics are the energy and the frequency spectrum [17]. The vibration or the cycle of opening/closing of the vocal cords represents f_0. This frequency characterizes only the voiced segments and evolves slowly over time [17]. The speaker frequency varies according to age and gender. It extends approximately from 60Hz to 150Hz for men and from 150Hz to 450Hz for women. The fundamental frequency extraction is not an easy task since the periodicity of vibration of the vocal cords is not always perfect [6]. The amplitude of the speech signal changes significantly over time and in the audible speech. This amplitude is much greater than in the inaudible speech and reflects to us the energy changes in the signal. One of the characteristics of the speech signal is the energy feature, usually symbolized by E, computed according to (2). This energy is very high in audible sounds compared to inaudible sounds in which the energy is weak [18].

$$E_X = \sum_n^{N-1}\left(X(n)\right)^2 \quad (1)$$

According to the scheme shown in Figure 1, we symbolize the standard called Wavelet Energy Rate (WER), which is applied to the first DWT level of the speech signal under study. This signal is a spoken sentence by some speakers of different genders containing repeated and different sentences. These are samples taken from a database that was prepared specifically for this study: videos of sentences in mp4 format were taken from people of both genders. These videos are downloaded and the spoken sentences were converted to wav format. Each sentence was sampled separately according to the sampling frequency fe =11025hz, which is a frequency greater than double the frequency of the speech signal (5khz) on the basis of Shannon's law.

III. FILTERING

The filtering step is accomplished through a mathematical transformation applied on the speech signal under study using a low pass filter with cutoff frequency $f_c = 600$Hz. This low pass filter reduces the signal high frequency to $-10\log_{10}$ (2)dB in the order of -3dB, which means decreasing the signal output energy to the 71% of the original speech signal [19]. Among the most widely used filters are the Butterworth linear filters, which are similar in shape with a difference in the cut-off frequency range. This filter has the largest amplitude and is more stable with frequency in the pass range and in the transition region with moderate reduction and is selected based on amplitude accuracy [20]. The filter order represents the number of columns in the filter pass region. For example, a filter of order n has a reduction rate of 6×n dB/decade to 20×n/decade. When n=8 its reduction is 48dB/ decade [20].

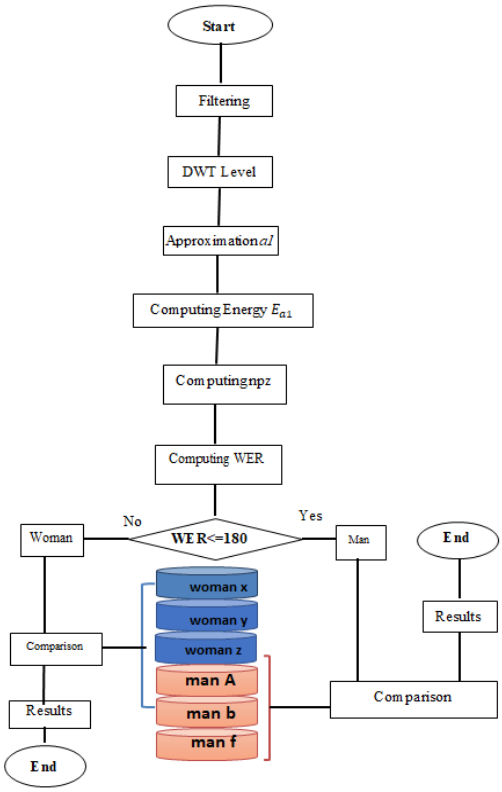

Fig. 1. General scheme.

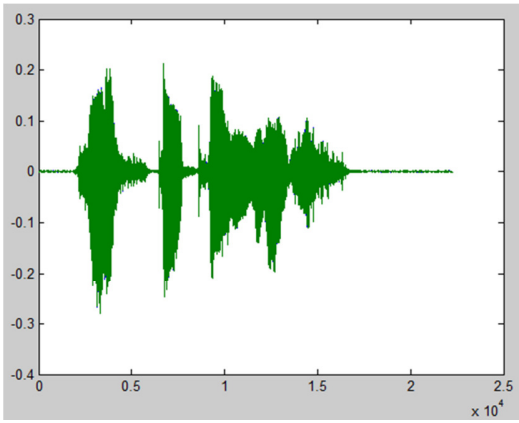

Fig. 2. Filtered speech signal.

IV. DISCRETE WAVELET TRANSFORM

DWT is based on sub-band coding, and developed after that to be a technique similar to sub-band coding called hierarchical coding [21]. It is identified by the following relationship [22]:

$$X_{DWT}[n] = x[n] * h[n] = \sum_{k=-\infty}^{\infty} x[k] * h[n-k] \quad (2)$$

DWT analyzes the original signal x(n) in different frequency bands with different degrees of accuracy by evaluating the signal into approximate and detailed coefficients where the approximate coefficients have the highest amplitude, the lowest frequencies, and possess most of the energy. The components of the highest frequencies are mixed with the noise existing in the original signal.

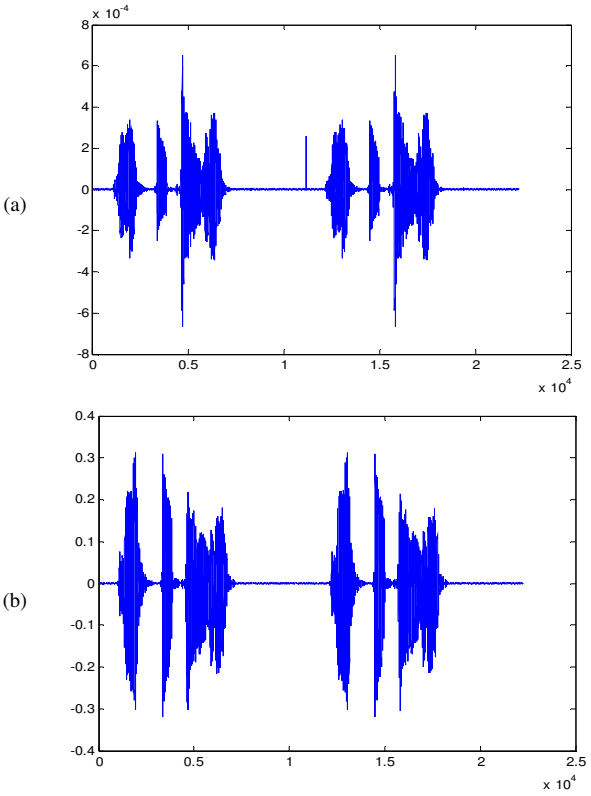

Fig. 3. First level signal DWT decomposition after filtering. (a) Approximation part *a1* of speech signal, (b) detail part of speech signal *d1*.

As previously mentioned, the DWT uses digital filtering techniques, in which the signal to be analyzed passes through filters with different cut-off frequencies at different scales. The original signal is divided into details coefficients and approximation coefficients, which are often denoted by ak and dk respectively, where d denotes the detailed coefficients and a denotes the approximate coefficients, while k denotes the analytical level. The two parameters are the result of the convolution of the original signal x(n) with the pass filter, meaning that the approximate parameters are the result of the low pass filter convolution with the original signal and the detailed parameters are the convolution of the high pass filter with the original signal [24].

$$a1 = \sum_{k=-\infty}^{\infty} X(k)y[2n-k] \quad (3)$$

$$d1 = \sum_{k=-\infty}^{\infty} X(k)z[2n-k] \quad (4)$$

In the waveform analysis, the signal is analyzed and synthesized according to stages starting from dividing the signal x(n) to the end of the required level and rebuilding it from the last level we stopped to the end of the first level from which we started. For example, if the original signal x(n) contains 512 samples and a frequency from 0 to π, then each of the approximate and detailed coefficients in the first level has 256 samples and one half(1/2) the frequency of the signal, i.e. π/2, and in the second level there are 128 samples and the half frequency of the signal level, i.e. f = π/4 or ¼ of the original signal frequency for both approximate and detailed parameters.

As mentioned above, the wavelet transform has more flexibility in designing the pulse shape and less sensitivity to signal distortion. The DWT analyzes the original signal x(n) in different frequency bands with different degrees of accuracy by analyzing the signal into approximate coefficients and detailed coefficients. The approximate coefficients have the highest amplitude and components of low frequencies and possess most of the signal energy. It is known that the speech signal is continuous and the fundamental frequency of a man's voice is around 50Hz (with period T = 20ms) whereas the f_0 of a woman's voice is near 180Hz, with a period T = 250ms. Also, the basic frequency for the voice of a woman and of a man rises to 500Hz and to 200Hz respectively.

V. INTERSECTIONS WITH ZERO

Intersections with zero is the number of the times the signal x(n) passes through zero in a certain period of time. It is an indication of the frequency at which the energy is concentrated in the signal spectrum, as the energy is concentrated at low frequencies in audible speech, while in inaudible speech most of it is found in high frequencies. From this, the highest number of intersections with zero is at high frequencies and the lowest is at low frequencies [18]. We can say that the intersection of the signal with zero has a strong relationship with the distribution of energy with frequency. The number of intersections with zero is computed for the signal according to [25]:

$$npz = \frac{1}{2N}\sum_{n=0}^{N-1}|sgn[x(n)] - sgn[x(n-1)]| \quad (5)$$

$$sgn[x(n)] = \begin{cases} +1 x(n) \geq 0 \\ -1 x(n) < 0 \end{cases} \quad (6)$$

VI. THE PROPOSED WAVELET ENERGY RATE (WER)

The WER criterion is calculated according to (7). This new criterion allows us to determine the gender of the speaker based on the studied values. WER is the quotient of the energy E_{a1} obtained from the approximation coefficient of the first level of the DWT applied on the signal x(n) according to (8) divided by npz:

$$WER = \frac{E_{a1}}{npz} \times Tpz \quad (7)$$

$$E_{a1} = \sum_{n}^{N-1}\left(a1(n)\right)^2 \quad (8)$$

where a_1 is the DWT first level, E_{a1} the DWT first signal approximation of coefficients energy, npz the DWT first level signal approximation of coefficient intersections with zero, and Tpz the intersection ratio with zero for the original signal.

The value of the WER criterion is distributed over a defined range from 50 to 180Hz for men and from 180 to 500Hz for women. These values are similar to the fundamental frequency of the speech signal f_0, which determines the gender of the speaker as a first process, before identifying the speaking person by comparison with other values in our previously prepared database.

VII. USED COMPUTATION TOOLS

In our investigation, we used KVideo Downloader4 program to download English learning videos in mp4 format

from an Internet Database. Then, these clips were converted to wav format using on line Covertio and on line Cloud Convert. The obtained signal was sampled into some specific sentences using the Audacity program by a sampling frequency fe=11025Hz to be analyzed and studied by the Matlab A13 toolbox on a compute with the following specifications: PC Acer based on x67, processor: Intel® Core™ i3-2348M, CPU: n2.30GHz, Version SMBIOS 2.7, Operating system: Windows 10

VIII. EXPERIMENTAL RESULTS

Through the obtained experimental results presented below, the developed approaches were applied on some different repeated sentences from male and female speakers. We run 200 experiments on the speech signal to extract about 4500 different values, after filtering the speech signal with a Butterworth filter of degree n=8 and sampling frequency fe=11025Hz. This signal was analyzed to the first level by DWT seventh partition (db7). The number of intersections with zero to be multiplied by the ratio of the original signal intersections with zero, according to (7) and (8) are shown in Table I.

Through Tables I and II, it was found that the energy of the original signal has different values in the two genders. The energy of the original signal for men is much greater than for women and this happens for the same repeated sentence, several different sentences for the same woman or the same man, or using the same sentence for several women and several men, as shown in Figure 9.

$$E_{SH} \gg E_{SF} \quad (9)$$

where E_{SH} is the energy of the original signal for men and E_{SF} the energy of the original signal for women.

TABLE I. A SENTENCE REPEATED FOR THE SAME PERSON AND FOR SEVERAL PEOPLE (4 WOMEN AND 3 MEN)

A sentence repeated for the same person and for several people (4 women and 3 men) *"do you speak english"*							
	Signal	**Sample number**	**Energy signal**	**Energy E_{a1}**	**ZCN signal**	ZCN_{a1}	ZCR_{a1}
Women	Woman A 601	22243	879380	28077000	1268	4405.3	0.0570
	Woman A 602	21060	833610	25851000	1224	4282.3	0.0581
	Woman B 201	12487	271940	9406900	740	2623.8	0.0592
	Woman B 203	12820	275960	10579000	811	2874.1	0.0632
	Woman D 301	12449	1795300	9581200	768	2701.7	0.0617
	Woman D 302	12793	2313000	10444000	815	2885.4	0.0637
	Woman M 1000	15493	577710	11501000	723	2600.5	0.0467
	woman M 1003	15988	583980	12943000	795	2833.3	0.0497
Men	Man B 101	13292	5851300	6844900	526	1790.1	0.0396
	Man B 102	13667	6016100	7346700	538	1864.2	0.0394
	Man K 100	11341	317560	4442000	391	1371.3	0.0345
	Man W 400	11685	773940	4254200	364	1285.6	0.0311
	Man W 401	11638	425690	4075200	354	1245.3	0.0304

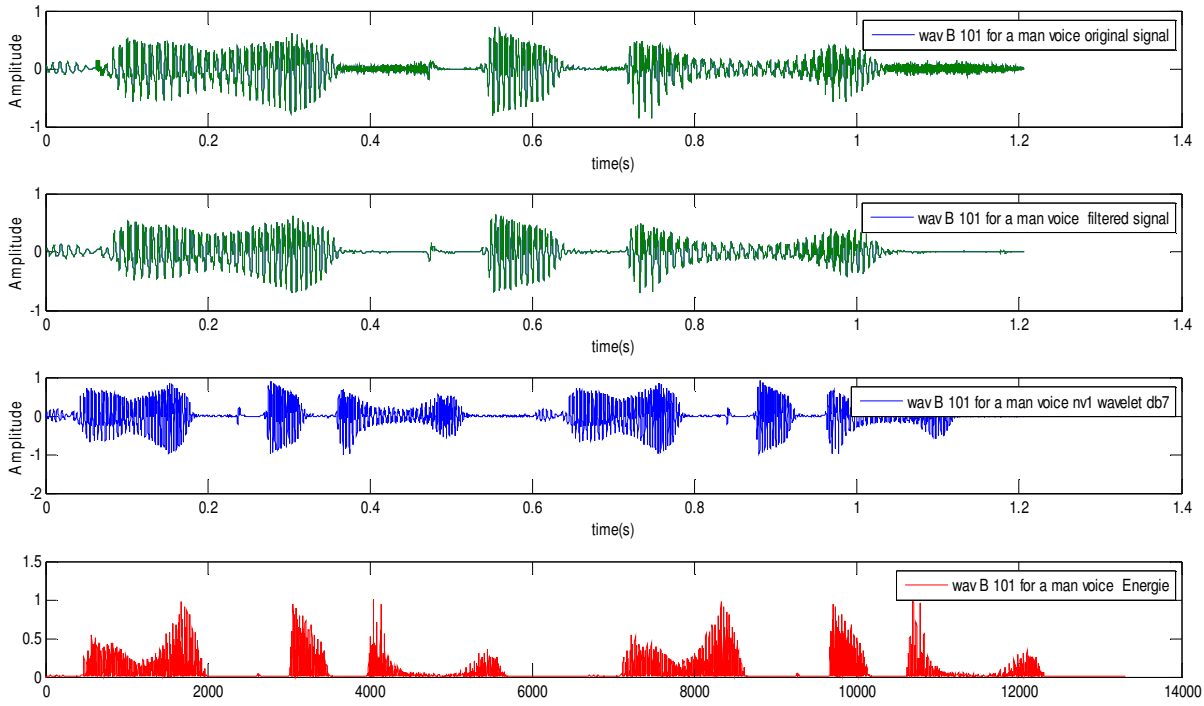

Fig. 4. Man B101 speech signal analysis (original, filtered, DWT decomposed, and energy spectrum), sentence: "do you speak English"?

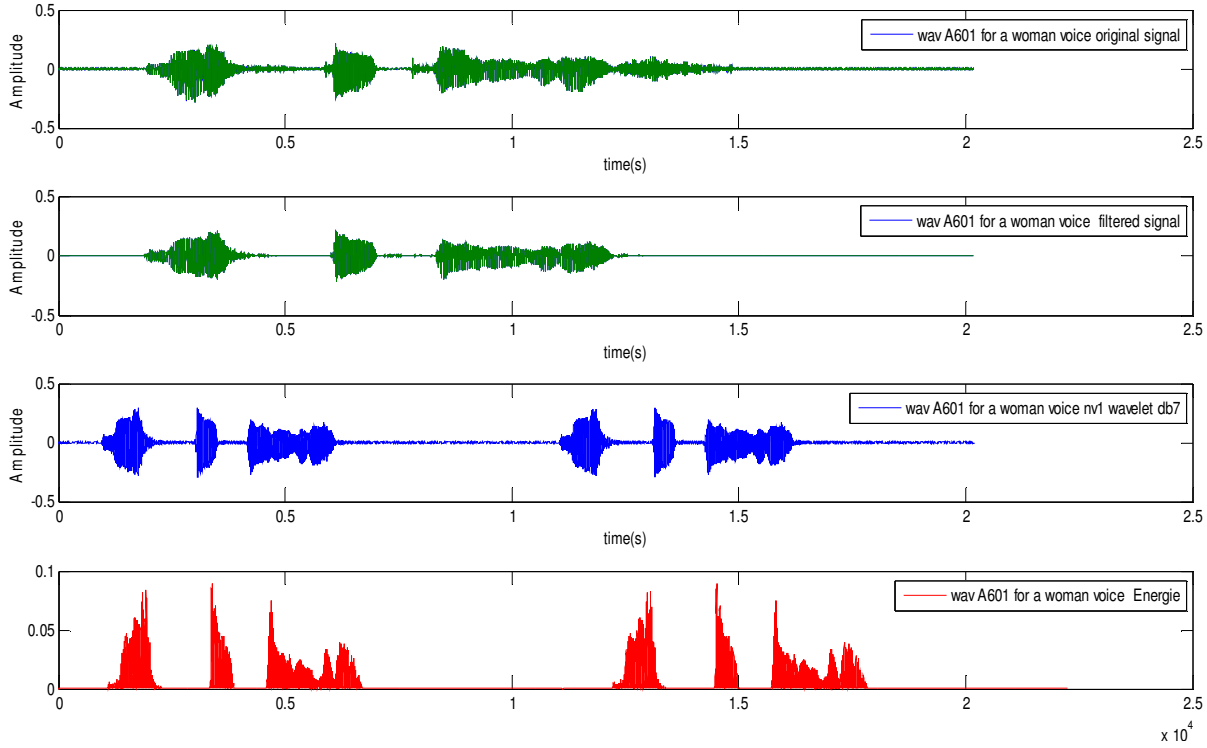

Fig. 5. Woman B601 speech signal analysis (original, filtered, DWT decomposed, and energy spectrum), sentence: "do you speak English"?

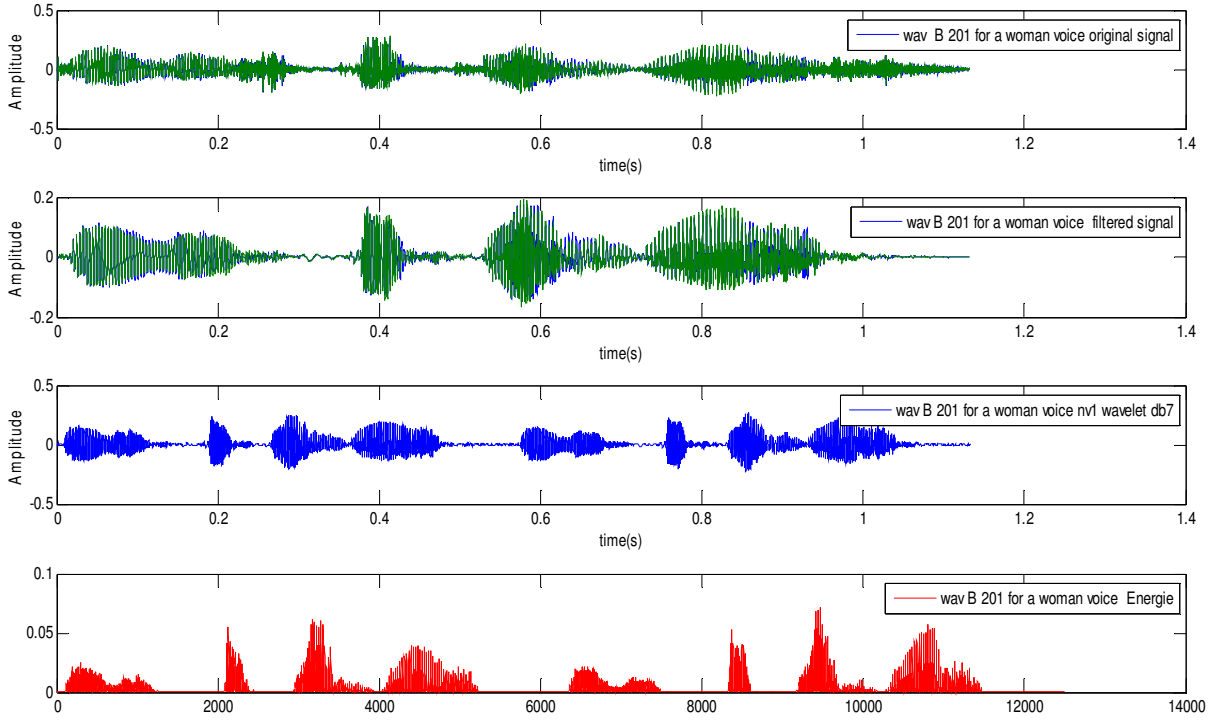

Fig. 6. Woman B201 speech signal analysis (original, filtered, DWT decomposed, and energy spectrum), sentence: "do you speak English"?

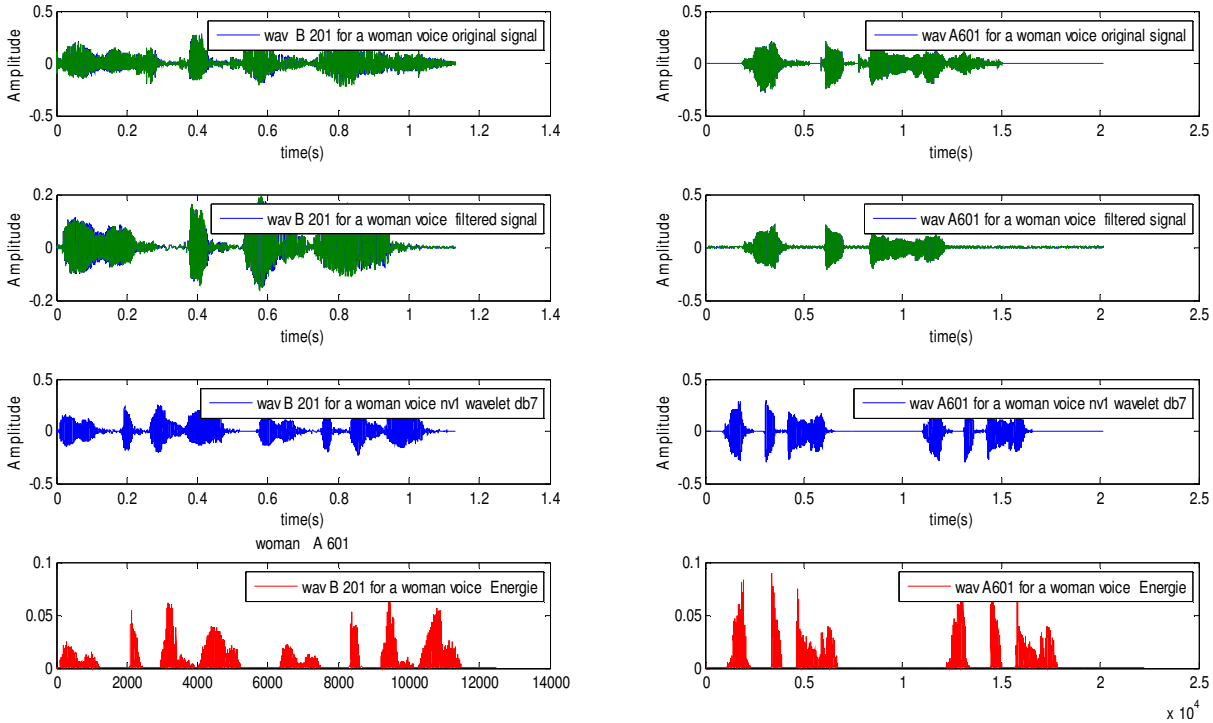

Fig. 7. Comparison between woman a601 and woman b201 speech signal analysis (original, filtered, DWT decomposed, and energy spectrum), sentence: "do you speak English"?

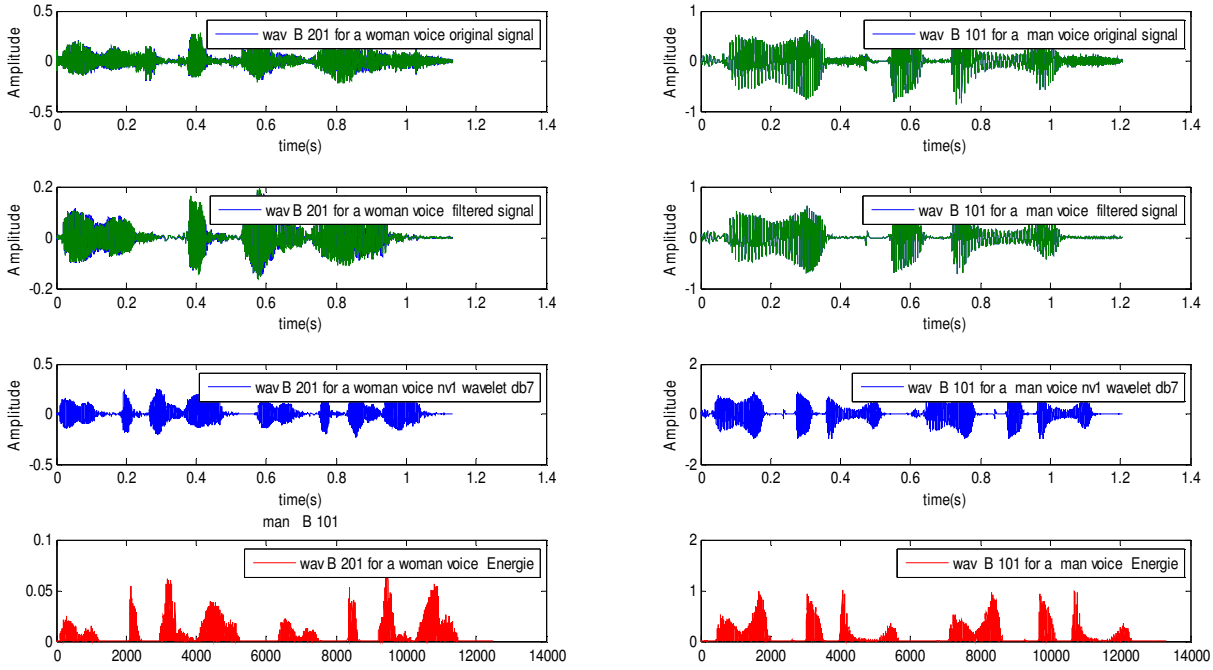

Fig. 8. Comparison between woman b201 and man b101 speech signal analysis (original, filtered, DWT decomposed, and energy spectrum), sentence: "do you speak English"?

In the approximate coefficient a_1 signal, in the first level of the DWT, we find that the energy Ea1 for women is greater than the energy for men and this is for the same repeated sentence or several different sentences for the same woman or the same man as shown in Tables I and II and Figure 10, or for a repetitive sentence or several different sentences for several women and several men:

$$E_{a1H} \gg E_{a1F} \quad (10)$$

where E_{a1H} E_{a1F} represent the energy of the approximate coefficient a1 for men and women respectively.

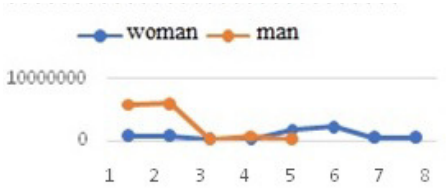

Fig. 9. Comparison of the signal energy between a man and a woman.

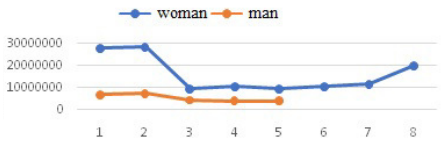

Fig. 10. Comparison of the Energy of the DWT first level signal between a man and a woman.

It is noticeable that the number of intersections with zero (Zero Crossing Number-ZCN) in the original signal differs for the two genders with diverse values. The ZCN of the original signal for men is much greater than for women, and for the same repeated sentence, several different sentences for the same woman or the same man, and for several different sentences and several men/women as shown in Tables I-II and Figure 11.

$$npz_{SF} \gg npz_{SH} \quad (11)$$

where npz_{SH} is the intersection of the original signal with zero for men and npz_{SF} for women.

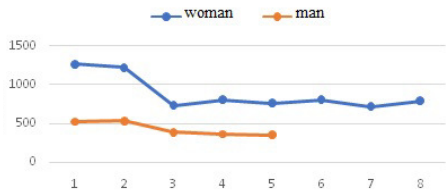

Fig. 11. ZCN comparison of for the signal between a man and a woman.

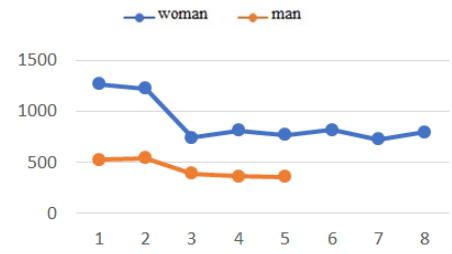

Fig. 12. ZCN comparison of in the DWT first level for the signals of a man and a moman.

On the other hand, for the approximate coefficient a1 signal in the first level of the DWT, we find that the zero crossing npz_{a1F} and the Zero Crossing Ratio (ZCR) for women is greater than that for men, for the same repeated sentence, several different sentences for the same woman, or for a repeated sentence or for different sentences for a number of men and women, as shown in Tables I-II and Figure 12. Through Figure13, we conclude that the crossing ratio with zero for the same level is greater for women than for men.

$$npz_{a1F} \gg npz_{a1H} \quad (12)$$

where npz_{a1H} and npz_{a1F} represent the crossing with zero of a1 signal for men and women respectively.

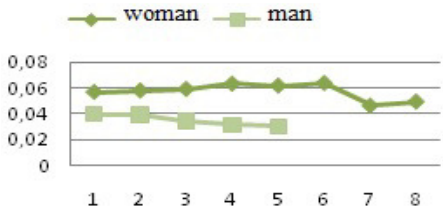

Fig. 13. Comparison of ZCR in the DWT first level for the signal of a man and a woman.

TABLE II. WER COMPARISON FOR MEN AND WOMEN

Signal	Sample number	Energy signal	Energy signal E_{a1}	ZCN_{a1}	ZCR_{a1}	WER
4 women and 2 men with the same sentence: *"do you speak English"?* repeated 3 times						
Woman A 601	22243	879380	28077000	4405.3	0.0570	363.3295
Woman A 602	21060	833610	25851000	4282.3	0.0581	350.8538
Woman B 201	12487	271940	9406900	2623.8	0.0592	212.4698
Woman B 202	12645	274830	9855600	2715.3	0.0606	219.8762
Woman D 301	12449	1795300	9581200	2701.7	0.0617	218.7851
Woman D 302	12793	2313000	10444000	2885.4	0.0637	230.6052
Woman m1000	15493	577710	11501000	2600.5	0.0467	206.3805
Woman m1003	15988	583980	12943000	2833.3	0.0497	227.1477
Man B 101	13292	5851300	6844900	1790.1	0.0396	151.3138
Man B 104	13666	6009400	7216300	1827.3	0.0391	154.3159
Man k 100	11341	317560	4442000	1373.3	0.0345	111.5123
Man w 400	11685	773940	4254200	1285.6	0.0311	103.0833
Man w 401	11638	425690	4075200	1245.3	0.0304	99.5563
4 women with the same sentence: *"see you later"*						
Woman A 2001	13595	420980	12289000	3129.3	0.0664	260.8490
Woman B 2003	9503	311870	6561400	2401.4	0.0721	197.2349
Woman D 402	5277	5571700	1748100	1171.4	0.0627	93.6040
Woman m 3002	10365	322280	5379100	1826.2	0.0506	149.1963
Man C201	12686	5501500	7000400	1923.6	0.0436	158.6357
Women and men with different sentences						
Woman A 401	20629	699960	24785000	4187.6	0.0583	345.1550
Woman A 901	12242	306010	9660000	2726	0.0649	230.1245
Woman A 803	9077	237420	5131100	1984.3	0.0622	160.9541
Woman A 202	13809	374400	12119000	3059.5	0.0632	250.4241
Woman A 101	12444	388450	10553000	2946.0	0.0676	242.3755
Woman B 402	12880	486000	8577500	2319.3	0.0497	183.7694
Woman B 501	8002	132600	5120800	2172.6	0.0813	191.7493
Woman B 802	20612	662150	23670000	3982.7	0.0552	328.4199
Woman B 901	11119	258280	7549900	2346.7	0.0620	199.6459
WomanB 3003	13112	247580	11385000	3006.3	0.0665	251.8563
Woman D 502	10428	2916500	7576900	2519.8	0.0696	209.3417
Woman D 202	11089	1078200	7591400	2390.4	0.0617	195.8928
Woman D 001	12320	1556200	9288400	2666.1	0.0613	213.5040
Man B 501	13160	2597500	8359000	2153.7	0.0487	189.0446
Man B 904	10492	2106500	3325500	1081.8	0.0299	92.0017
Man B 144	11025	2578600	5503800	1718.6	0.0455	145.8184
Man B1004	10227	2481400	4480800	1520.2	0.0434	127.9606
Man B193	10998	2365600	5382100	1678.9	0.0446	143.1177
Man X102	8405	391600	5291400	2182.7	0.0748	181.4226
Man X 100	14152	1428800	13458000	3296.2	0.0668	272.9188
Man C100	11710	2518300	3975700	1196.7	0.0292	97.0313

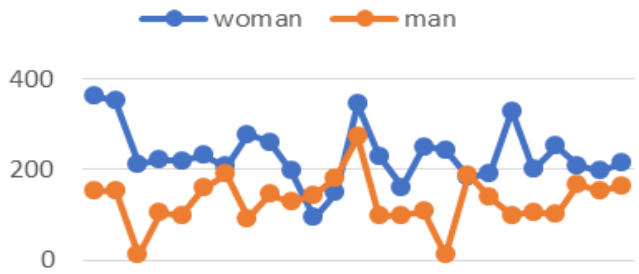

Fig. 14. Comparison of WER for different speakers of both sexes.

The values shown in Table II give us the WER values for the studied speech signals x(n). The signal is a repeated sentence for the same woman or the same man, the same sentence for several men and several women, or a speech signal containing different sentences for several men and several women, as shown in Figure 14. A comparison of these values in relation to the range of the fundamental frequency f_0 values in recently published research works was made. The result is that the WER criterion provides values approximately close to the values of f_0 for speakers of different sentences and gender. Accordingly, we can through WER specify values in a range from 50 to 180Hz for male speakers and values in a range from 180 to 504Hz for female speakers, which are almost similar to the range of fundamental frequency values which determine the speaker gender according to the value of his/her fundamental frequency using different methods. The direct computational speed of the WER is a faster and easier way to determine the gender of a speaker. Table II illustrates the comparison between WER values and Figure 14 shows their progress.

After this stage, which determines the gender of the speaker, the next stage of verifying the speaker is followed by reference to our stored database, according to the comparison process based on a suitable algorithm to authenticate and confirm the speaker as shown by the flowchart in the general scheme of Figure 1.

IX. CONCLUSION

In speech investigation, gender identification is usually performed by extracting the information from the speech signals. In this paper, we propose a novel criterion for determining the speaker gender by defining a range of sampled values using the Discrete Wavelet Transform (DWT), and computing the energy in addition to the intersections with zero (ZCR) of the first level DWT followed by estimating the introduced criterion (WER) values for both genders. Comparison was made with the value of the speech fundamental frequency for speakers of both sexes. The proposed criterion was tested on a large database prepared for this research, containing many repeated sentences from the same person and from persons of both sexes and gives acceptable results, proving its suitability for speaker gender identification.

ACKNOWLEDGMENT

The authors would like to thank the LAADI Research Laboratory for the technical support.

REFERENCES

[1] L. Jeancolas, "Détection précoce de la maladie de Parkinson par l'analyse de la voix et corrélations avec la neuroimagerie," Ph.D. dissertation, Paris-Saclay University, Paris, France, 2019.

[2] R. Ajgou, "Techniques De Détection De La Période Du Pitch Par Les Méthodes Temps Fréquence Et Temps Échelle.," M.S. thesis, University of Biskra, Biskra, Algeria, 2010.

[3] F. Bahja, *Détection du fondamental de la parole en temps-réel: Application aux voix pathologiques*. Presses Académiques Francophones, 2014.

[4] R. Ajgou, S. Sbaa, S. Aouragh, and A. Taleb, "Détection Du Pitch Par Les Ondelettes Continues En Temps Réel Pour Un Signal Parole Basée Sur Un Seuil Adaptatif Pour Une Détermination V/Nv," *Courrier du Savoir Scientifique et Technique*, vol. 12, no. 12, pp. 21–26, May 2014.

[5] M. A. Ben Messaoud, A. Bouzid, and N. Ellouze, "Estimation du pitch et décision de voisement par compression spectrale de l'autocorrélation du produit multi-échelle (Pitch estimation and voiced decision by spectral autocorrelation compression of multi-scale product) [in French]," in *Proceedings of the Joint Conference JEP-TALN-RECITAL 2012*, Grenoble, France, Mar. 2012, vol. 1, pp. 201–208.

[6] Y. Fayçal, R. Amiar, S. Hecini, W. Benzaba, and L. Bendaouia, "Etude Comparative des Performances de Plusieurs Techniques de Détection de la Fréquence Fondamentale des Signaux Vocaux.," in *Proceedings of the 2nd Conférence Internationale sur l'Informatique et ses Applications (CIIA'09)*, Saida, Algeria, Jan. 2009.

[7] M. A. Nasr, M. Abd-Elnaby, A. S. El-Fishawy, S. El-Rabaie, and F. E. Abd El-Samie, "Speaker identification based on normalized pitch frequency and Mel Frequency Cepstral Coefficients," *International Journal of Speech Technology*, vol. 21, no. 4, pp. 941–951, Dec. 2018, https://doi.org/10.1007/s10772-018-9524-7.

[8] M. Chandra, P. Nandi, A. kumari, and S. Mishra, "Spectral-Subtraction Based Features for Speaker Identification," in *Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA)*, 2015, pp. 529–536, https://doi.org/10.1007/978-3-319-12012-6_58.

[9] S. R. Shahamiri and F. Thabtah, "An investigation towards speaker identification using a single-sound-frame," *Multimedia Tools and Applications*, vol. 79, no. 41, pp. 31265–31281, Nov. 2020, https://doi.org/10.1007/s11042-020-09580-4.

[10] I. Vélez, C. Rascon, and G. Fuentes-Pineda, "Lightweight speaker verification for online identification of new speakers with short segments," *Applied Soft Computing*, vol. 95, Oct. 2020, Art. no. 106704, https://doi.org/10.1016/j.asoc.2020.106704.

[11] W. Helali, Z. Hajaiej, and A. Cherif, "Real Time Speech Recognition based on PWP Thresholding and MFCC using SVM," *Engineering, Technology & Applied Science Research*, vol. 10, no. 5, pp. 6204–6208, Oct. 2020, https://doi.org/10.48084/etasr.3759.

[12] K. Daqrouq and K. Y. Al Azzawi, "Average framing linear prediction coding with wavelet transform for text-independent speaker identification system," *Computers & Electrical Engineering*, vol. 38, no. 6, pp. 1467–1479, Nov. 2012, https://doi.org/10.1016/j.compeleceng.2012.04.014.

[13] C. Turner and A. Joseph, "A Wavelet Packet and Mel-Frequency Cepstral Coefficients-Based Feature Extraction Method for Speaker Identification," *Procedia Computer Science*, vol. 61, pp. 416–421, Jan. 2015, https://doi.org/10.1016/j.procs.2015.09.177.

[14] M. A. Nasr, M. Abd-Elnaby, A. S. El-Fishawy, S. El-Rabaie, and F. E. Abd El-Samie, "Speaker identification based on normalized pitch frequency and Mel Frequency Cepstral Coefficients," *International Journal of Speech Technology*, vol. 21, no. 4, pp. 941–951, Dec. 2018, https://doi.org/10.1007/s10772-018-9524-7.

[15] M. Kiran Reddy *et al.*, "The automatic detection of heart failure using speech signals," *Computer Speech & Language*, vol. 69, Sep. 2021, Art. no. 101205, https://doi.org/10.1016/j.csl.2021.101205.

[16] A. Mnassri, M. Bennasr, and C. Adnane, "A Robust Feature Extraction Method for Real-Time Speech Recognition System on a Raspberry Pi 3 Board," *Engineering, Technology & Applied Science Research*, vol. 9, no. 2, pp. 4066–4070, Apr. 2019, https://doi.org/10.48084/etasr.2533.

[17] A. Amehraye and S. Saoudi, *Débruitage perceptuel de la parole*. 2009.

[18] R. Narayanam, "Voiced and Unvoiced Separation in Speech Auditory Brainstem Responses of Human Subjects Using Zero Crossing Rate (ZCR) and Energy of the Speech Signal," *International Journal of Engineering Sciences & Research Technology*, vol. 4, no. 9, pp. 370–380, Jun. 2017, https://doi.org/10.5281/zenodo.892088.

[19] "Fréquence de coupure," *Wikipédia*. Feb. 11, 2022, [Online]. Available: https://fr.wikipedia.org/w/index.php?title=Fr%C3%A9quence_de_coupure&oldid=190757368.

[20] M. V. Daithankar and S. D. Ruikar, "Analysis of the Wavelet Domain Filtering Approach for Video Super-Resolution," *Engineering, Technology & Applied Science Research*, vol. 11, no. 4, pp. 7477–7482, Aug. 2021, https://doi.org/10.48084/etasr.4262.

[21] A. Pini, "Notions de base sur les filtres passe-bas antirepliement (et pourquoi ils doivent être adaptés au CAN)," *Digi-Key Electronics*, Mar. 24, 2020. https://www.digikey.fr/fr/articles/the-basics-of-anti-aliasing-low-pass-filters.

[22] D. Sripath, "Efficient Implementations of Discrete Wavelet Transforms Using FPGAs," Jan. 2003.

[23] E. Hostalkova, "Wavelet Transform," Athens, Greece, Nov. 2009.

[24] A. Sumithra and B. Thanushkodi, "Performance Evaluation of Different Thresholding Methods in Time Adaptive Wavelet Based Speech Enhancement," *International Journal of Engineering and Technology*, vol. 1, no. 5, pp. 439–447, 2009, https://doi.org/10.7763/IJET.2009.V1.82.

[25] K. Tajane, R. Pitale, and J. Umale, "Review Paper :Comparative Analysis Of Mother Wavelet Functions With The ECG Signals," *International Journal of Engineering Research and Applications*, vol. 4, no. 1, pp. 38–41, Jan. 2014.