entropy

MDPI

Article

Noise Suppression in 94 GHz Radar-Detected Speech Based on Perceptual Wavelet Packet

Fuming Chen [1,†], Chuantao Li [1,2,†], Qiang An [1], Fulai Liang [1], Fugui Qi [1], Sheng Li [3,*] and Jianqi Wang [1,4,*]

1 Department of Biomedical Engineering, Fourth Military Medical University, Xi'an 710032, China; cfm5762@126.com (F.C.); lichuantao614@126.com (C.L.); anqiang900903@163.com (Q.A.); liangfulai@fmmu.edu.cn (F.L.); qifgbme@outlook.com (F.Q.)

2 Laboratory of Aviation Medicine, Naval Medical Research Institute, Shanghai 200433, China

3 College of Control Engineering, Xijing University, Xi'an 710123, China

4 Department of Physics and Information Engineering, Shaanxi University of Technology, Hanzhong 723001, China

* Correspondence: sheng@mail.xjtu.edu.cn (S.L.); wangjq@fmmu.edu.cn (J.W.); Tel.: +86-29-8477-4843 (J.W.); Fax: +86-29-8477-9259 (J.W.)

† These authors contributed equally to this work.

Academic Editor: Carlo Cattani

Received: 12 April 2016; Accepted: 13 July 2016; Published: 19 July 2016

Abstract: A millimeter wave (MMW) radar sensor is employed in our laboratory to detect human speech because it provides a new non-contact speech acquisition method that is suitable for various applications. However, the speech detected by the radar sensor is often degraded by combined noise. This paper proposes a new perceptual wavelet packet method that is able to enhance the speech acquired using a 94 GHz MMW radar system by suppressing the noise. The process is as follows. First, the radar speech signal is decomposed using a perceptual wavelet packet. Then, an adaptive wavelet threshold and new modified thresholding function are employed to remove the noise from the detected speech. The results obtained from the speech spectrograms, listening tests and objective evaluation show that the new method significantly improves the performance of the detected speech.

Keywords: radar-detected speech; 94 GHz MMW radar; speech enhancement; perceptual wavelet packet; thresholding function

1. Introduction

Speech signals carry a great deal of information that is essential for effective human communication. It is well known that speech can be transmitted through air and can be detected by traditional acoustic transducers, or air-borne microphones that convert acoustic energy into electrical energy [1]. Other methods to detect speech signals include using bone conduction microphones, which are transducers that detect vibrations conducted through bone [2,3], and optical techniques, such as light waves or lasers [4]. While all of these methods are commonly used, they do have some potential limitations. The directional sensitivity of microphones is quite weak and can be easily disturbed by ambient noise. Bone conduction microphones need to be applied to the throat with adhesive, which may cause discomfort to some users. Optical speech signals are strongly affected by environmental conditions, including atmospheric conditions, the composition of the target, and the properties of the objects [5]. In addition, the details of the optical materials in use are often difficult to obtain [6]. To overcome these limitations, microwave radar speech sensors have been employed to detect the motion of vocal cords [7–9]; however, the research into these sensors has been concentrated primarily on the relationship between the motion of the organ and the voice activity, and has seldom been dedicated to studying the speech signal itself.

Microwave radar has many advantages, such as low range attenuation and a good sense of direction, and is noninvasive, safe, fast, and portable. For these reasons, microwave radar systems have been developed for use in a variety of remote sensing applications [10–12]. Millimeter wave (MMW) radar can generate a larger modulated phase and has a higher sensitivity to small vibration displacements than does centimeter wave radar. Based on these advantages, the use of MMW radar in speech-detecting applications will provide many exciting possibilities in the future. In [6], Li verified the feasibility of this method by using a 40 GHz MMW radar to acquire speech signals with a 40 GHz MMW radar and verified the feasibility of this method. However, no other examples have been found in the literature. We therefore developed a 34 GHz microwave radar sensor for non-contact speech detection [13,14]; however, the detection sensitivity was low and the quality of the detected speech was deemed to be unsatisfactory. We believe it is important to explore this new method to detect speech signals, but first there is an urgent need to improve the detection sensitivity of the radar sensor.

Recently, Obeid et al. used three measurement systems to detect non-contact heartbeats and determined that a high operating frequency can increase the sensitivity to small displacements [15]. Mikhelson employed a 94 GHz radar to detect smaller displacements and successfully acquired signals that contain both the heart rate and respiration patterns of a human subject [16]. These studies suggest that systems that use a high operating frequency will demonstrate a high detection sensitivity. Therefore, to improve the sensitivity of a radar sensor used to detect speech, we have employed a higher frequency radar sensor that operates at 94 GHz in our laboratory [17]. To allay user concerns, the radar poses no risk to human health according to a new standard for safety levels with respect to human exposure to radio-frequency radiation [18].

Although the 94 GHz radar sensor performed better than the sensor at 34 GHz when detecting speech, it also had several serious shortcomings, including reduced intelligibility and poor audibility [17,19]. This is because the quality of the detected speech was significantly degraded by several sources of noise, which include electromagnetic, ambient, and channel noise. These noise sources are more complex than those found in speech acquired using a traditional microphone. Therefore, a challenging task for researchers is to determine how best to reduce the level of this combined noise in order to enhance the quality of the detected speech. Traditionally, various speech enhancement approaches have been proposed to solve this problem, such as Wiener filtering and spectral subtraction [20–23]. Our laboratory has also proposed some noise reduction methods to enhance the quality of the speech detected by the radar sensor [24,25].

Wavelet shrinkage is a simple denoising technique based on the thresholding of the wavelet coefficients. This method was introduced by Donoho as a powerful tool for denoising signals degraded by additive white noise [26–30], and has now been widely applied. For example, Chambolle et al. used the method for image denoising and compression [31]. Achim et al. proposed a new synthetic aperture radar (SAR) image denoising method via Bayesian wavelet shrinkage [32]. Bahoura et al. proposed a new speech enhancement method based on the time adaption of wavelet shrinkage that was able to successfully reduced the noise in speech signals [33]. Although the principle of wavelet shrinkage for denoising signals has now been applied in many different areas, the method does require several assumptions that limit the algorithm itself [30]. To address this, Mercorelli proposed a thresholding-free methodology that has benefits in wavelet denoising applications [34–36]. However, there are still two important issues with the wavelet shrinkage method that need to be improved. First, the down sampling at each level of decomposition, leads to speech signal distortion and a high computational load [37]. Second, the disadvantages of the hard and soft thresholding functions result in an unsatisfactory quality of the reconstructed speech signals.

This paper introduces a non-contact MMW radar system for acquiring high quality speech signals that addresses these problems by proposing a perceptual wavelet packet method that is able to enhance the perceptibility and intelligibility of radar-detected speech in two specific ways. First, the new method offers better resolution when decomposing signals because it allows the wavelet tree to be adjusted based on the critical bands. Second, a modified thresholding function is employed to

reduce the noise in the radar-detected speech. Experimental results demonstrate that the proposed algorithm is an effective way to suppress the noise in the speech detected by the 94 GHz radar system.

2. The 94 GHz MMW Radar System

A block diagram of the 94 GHz MMW radar system is shown in Figure 1. A dual-antenna structure is used and two Cassegrain antennas, each with a diameter of 200 mm, are used to focus the beam on the target. The gain of the antennas is 41.7 dBi, the beam width is 1° at −3 dB levels, and the operating wavelength is 3.19 mm. The system uses a separate transmitter and receiver in order to improve the isolation between the transmit and receive antennas. The output radio frequency (RF) power of the transmit antenna is 100 mW. The waveguide band WR-10 was selected because it provides small propagation losses over long distances when monitoring subtle physiological movements. The operating principle of the radar is described as follows. A continuous wave signal is generated by the Dielectric Resonator Oscillator (DRO) at 7.23 GHz, and is then amplified and fed into both the transmitter and receiver modules. In the transmitter module, the signal is processed by a frequency multiplier and bandpass filter before being radiated by the antenna toward the throat of the human subject. In the receiver module, the reflected signal is received by the receive antenna, and is then balance-mixed with a locally processed signal at a frequency of 86.7 GHz. Next, the processed signal is amplified by a low-noise amplifier (LNA) and is then mixed with two quadrature local signals in the in-phase and quadrature (I/Q) mixer. After I/Q quadrature demodulation, the signal is sampled by an A/D converter before being transferred to a computer. Finally, the signal is sent to a D/A converter and then passed through a power amplifier (PA) in order to drive a speaker.

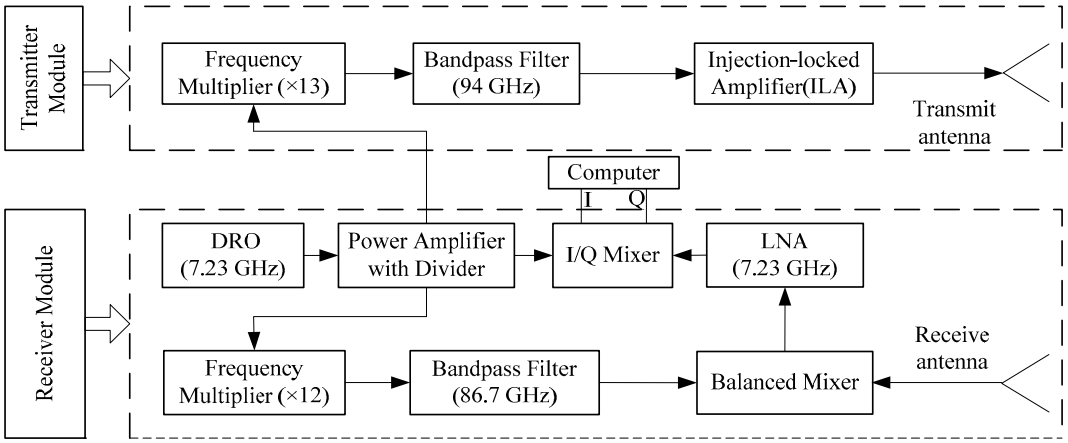

Figure 1. Schematic diagram of the 94 GHz MMW radar system.

3. Signal Recording and Processing

Four volunteers (three males and one female) were selected for the speech detection experiments, and all were Chinese native speakers. The radar speech sensor was positioned at distances ranging from 2 to 10 m from the subjects. In the experiment, a volunteer sat in front of the radar system with his throat at the same height as the radar antenna. The speech material selected for the volunteers consisted of two Mandarin Chinese sentences and the standard library from the TIMIT Database [38]. In order to acquire high quality and stable speech, a loudspeaker was also used to represent the speakers and play the speech material. All of the experimental procedures were in accordance with the rules of the Declaration of Helsinki [39].

After acquiring the radar-detected speech signals, a series of signal processing methods were employed to suppress the noise in the detected speech, as shown in Figure 2. First, the detected speech was decomposed into a bark band tree using the perceptual wavelet packet method described in Section 3.2. Then, an adaptive wavelet threshold and modified thresholding function were employed

to remove the noise from the detected speech, as detailed in Section 3.3. Finally, the processed signal was used to reconstruct the speech that was detected using the radar system.

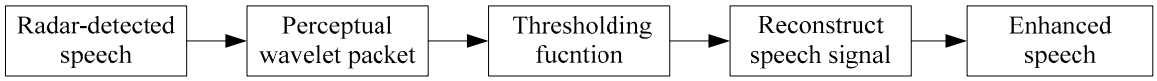

Figure 2. Block diagram of the signal processing method.

3.1. Wavelet Packet Transform

The wavelet packet transform is an extension of the wavelet transform and was pioneered by Coifman et al. [40,41]. In contrast to the wavelet transform, the wavelet packet transform decomposes both the low frequencies and the high frequencies. It can therefore offer better resolution than the wavelet transform.

If the quadrature mirror filter coefficients of the wavelet multi-resolution analysis are $h(k)$ and $g(k)$, then the wavelet basic function $\varphi(t)$ and its corresponding orthonormal scaling function ψ(t) are given by:

$$\begin{cases} \varphi(t) = \sum\limits_{k\in Z} h(k)\varphi(2t-k) \\ \psi(t) = \sum\limits_{k\in Z} g(k)\varphi(2t-k) \end{cases} \tag{1}$$

The function cluster of $u_n(t)$ is defined to satisfy the following equation:

$$\begin{cases} u_{2n}(t) = \sqrt{2}\sum\limits_{k\in Z} h(k)u_n(2t-k) \\ u_{2n+1}(t) = \sqrt{2}\sum\limits_{k\in Z} g(k)u_n(2t-k) \end{cases} \tag{2}$$

where $g(k) = (-1)^k h(1-k)$. When $n = 0$, the equation can be expressed as:

$$\begin{cases} u_0(t) = \sqrt{2}\sum\limits_{k\in Z} h(k)u_0(2t-k) \\ u_1(t) = \sqrt{2}\sum\limits_{k\in Z} g(k)u_0(2t-k) \end{cases} \tag{3}$$

where $u_0(t)$ can be identified with the function $\varphi(t)$, and $u_1(t)$ with the function $\psi(t)$. Then, the collection function family $\{u_n(t)\,|n \in Z^+\}$ can be described as wavelet packets obtained from the function $\varphi(t)$.

For a radar-detected speech signal, if the noisy signal is represented by $y(n)$, the clean signal by $x(n)$, and the uncorrelated additive noise signal by $d(n)$, then:

$$y(n) = x(n) + d(n) \qquad n = 0,\ 1\cdots,\ N-1 \tag{4}$$

where N is the number of samples in the received radar-detected speech signal.

An approximation of an original noisy speech signal function $y(n)$ using wavelet packets can be written as $y_j^n(n) \in U_j^n$,

$$y_j^n(n) = \sum_l d_l^{j,n} u_n(2^j n - l) \tag{5}$$

The wavelet packet decomposes the noisy signal $y(n)$ into 2^j subbands, with the corresponding wavelet coefficient sets as $d_l^{j,n}$, and it denotes the nth coefficient of j-th subband for the l-th level, and $n = 1,\cdots,(N/2^j) - 1$, $j = 1,\cdots,2^l$, where j, n, and l can be regarded as the scale, frequency, and position indices of the corresponding wavelet packet function, respectively.

Then, the wavelet packet transform can be constructed using basic decomposition and reconstruction techniques. The wavelet packet decomposition can be given by:

$$\left.\begin{aligned} d_k^{j+1,2n} &= \sum_k h_{2l-k} d_l^{j,n} \\ d_k^{j+1,2n+1} &= \sum_k g_{2l-k} d_l^{j,n} \end{aligned}\right\} \tag{6}$$

where ${d_k}^{j+1,2n}$ and ${d_k}^{j+1,2n+1}$ are called the approximation coefficients and the detail coefficients of the wavelet decomposition of ${d_l}^{j,n}$, respectively, and h_{2l-k} and g_{2l-k} are the analysis low-pass scaling filter and the high-pass wavelet filter, respectively [42].

After the coefficients of ${d_k}^{j+1,2n}$ and ${d_k}^{j+1,2n+1}$ have been processed by the wavelet threshold, we obtain the updated wavelet packet coefficients $\hat{d}_k^{j+1,2n}$ and $\hat{d}_k^{j+1,2n+1}$. The enhanced speech is then synthesized with the inverse transformation of the processed wavelet packet coefficients, and the corresponding wavelet reconstruction can be written as:

$$\hat{d}_l^{j,n} = \sum_k [h_{l-2k} \hat{d}_k^{j+1,2n} + g_{l-2k} \hat{d}_k^{j+1,2n+1}] \tag{7}$$

where h_{k-2l} and g_{k-2l} are the synthesis low-pass scaling filter and the high-pass wavelet filter, respectively. Then, the enhanced radar-detected speech signal can be obtained using Equation (8)

$$\hat{y}_j^n(n) = \sum_l \hat{d}_l^{j,n} u_n(2^j n - l) \tag{8}$$

3.2. Perceptual Wavelet Packet Decomposition

Theoretically, the human auditory frequency range extends from 0 to 16 kHz. The Bark scale is a human auditory scale that was proposed by Zwicker [43] and divides the human auditory range into approximately 24 critical bands [43]. The perceptual wavelet packet decomposition method described in this paper is used to decompose the speech signal from 0 Hz to 16 kHz into 24 frequency subbands that approximate the 24 Bark bands. The Bark $z(f)$ band can be approximately described as the relationship between the linear frequency and the critical band number [44]:

$$z(f) = 13\tan^{-1}(7.6 \times 10^{-4} f) + 3.5\tan^{-1}(1.33 \times 10^{-4} f)^2 \text{ [Bark]} \tag{9}$$

where f is the linear frequency in Hertz. The corresponding critical bandwidth (CBW) of the center frequencies can be expressed by [44]:

$$\text{CBW}(f_c) = 25 + 75(1 + 1.4 \times 10^{-6} {f_c}^2)^{0.69} \text{[Hz]} \tag{10}$$

where f_c is the center frequency in Hertz.

In this paper, the frequency range of the radar-detected speech is assumed to extend from 0 to 4000 Hz. According to Equation (9), there are approximately 17 critical bands. The tree structure of the perceptual wavelet packet can be decomposed into five levels, as shown in Figure 3.

Table 1 shows the specifications for the center frequencies (f_{center}), CBW, lower cutoff frequencies (f_l) and upper cutoff frequencies (f_u) for the 17 critical bands and the perceptual wavelet packet tree. Figure 4a shows the center frequencies of the critical bands and the perceptual wavelet packet tree. Figure 4b shows the CBW of the critical bands and the perceptual wavelet packet tree. It can be seen from Table 1 and Figure 4 that the perceptual wavelet packet tree closely approximates the critical bands of the human auditory system. This demonstrates that the proposed perceptual wavelet packet method decomposes the detected speech signal $y(n)$ into 17 subbands that correspond to wavelet coefficient sets ${d_l}^{j,n}$, where, l = 3, 4, 5 and j = 1,$\cdots$, 17. It should be noted that better results can be

achieved when suppressing the noise in radar-detected speech if the wavelet basis is chosen to be Daubechies6 (db6).

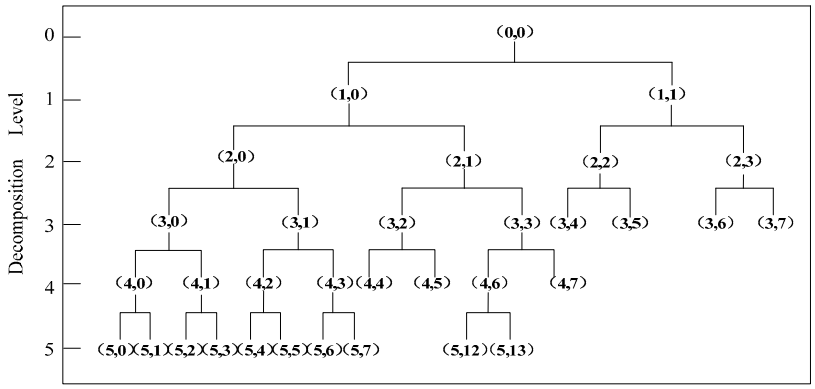

Figure 3. Tree structure of the perceptual wavelet packet decomposition.

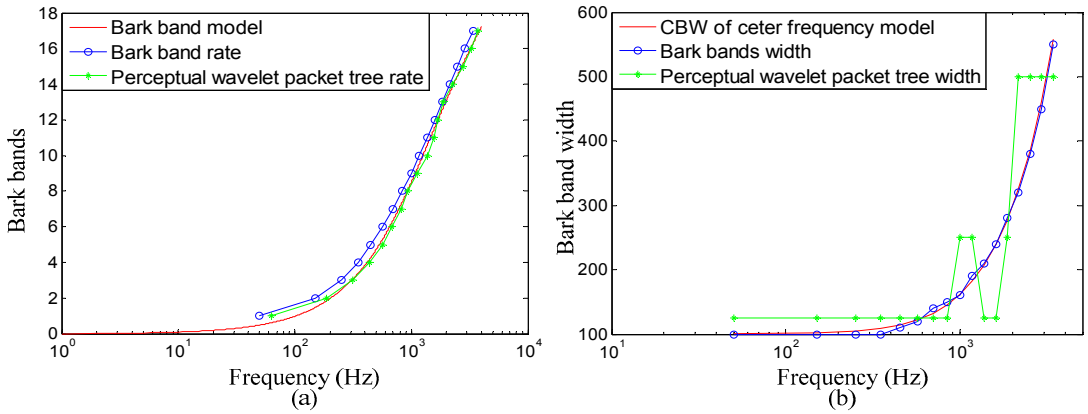

Figure 4. (**a**) Center frequency of the critical bands and the perceptual wavelet packet tree; (**b**) CBW of the critical bands and the perceptual wavelet packet tree.

Table 1. Comparison of the 17 critical bands and the perceptual wavelet packet tree.

z	Critical Bands			Perceptual Wavelet Tree		
	f_{center}	$[f_l\ f_u]$	CBW	f_{center}	$[f_l\ f_u]$	CBW
1	50	[0 100]	100	62.5	[0 125]	125
2	150	[100 200]	100	187.5	[125 250]	125
3	250	[200 300]	100	312.5	[250 375]	125
4	350	[300 400]	100	437.5	[375 500]	125
5	450	[400 510]	110	562.5	[500 625]	125
6	570	[510 630]	120	687.5	[625 750]	125
7	700	[630 770]	140	812.5	[750 875]	125
8	840	[770 920]	150	937.5	[875 1000]	125
9	1000	[920 1080]	160	1125	[1000 1250]	250
10	1170	[1080 1270]	190	1375	[1250 1500]	250
11	1370	[1270 1480]	210	1562.5	[1500 1625]	125
12	1600	[1480 1720]	240	1687.5	[1625 1750]	125
13	1850	[1720 2000]	280	1875	[1750 2000]	250
14	2150	[2000 2320]	320	2250	[2000 2500]	500
15	2500	[2320 2700]	380	2750	[2500 3000]	500
16	2900	[2700 3150]	450	3250	[3000 3500]	500
17	3400	[3150 3700]	550	3750	[3500 4000]	500

3.3. Modified Thresholding Function

The wavelet threshold plays an important role in suppressing the noise in radar-detected speech. Donoho has presented a very concise method to estimate the wavelet coefficients $d_i(n)$. The threshold of Donoho was given by [29]:

$$T_j = \sigma_j \sqrt{2\ln(N_j)} \tag{11}$$

where N_j is the signal length in scale j, and σ is the estimated noise level and is defined by:

$$\sigma_j = \mathrm{MAD}_j / 0.675 \tag{12}$$

where MADj is the median absolute deviation estimated on scale j.

The wavelet thresholding function is also an important factor in wavelet-based methods. The standard thresholding functions used in the wavelet shrinkage algorithm are the soft and hard thresholding functions. The soft thresholding function can be written as:

$$T(n) = \begin{cases} sign\{d_i(n)\}\{d_i(t) - T_j\}, & if \ |d_i(n)| > T_j \\ 0, & if \ |d_i(n)| \leqslant T_j \end{cases} \tag{13}$$

The hard thresholding function can be written as:

$$T(n) = \begin{cases} d_i, & if \ |d_i(n)| > T_j \\ 0, & if \ |d_i(n)| \leqslant T_j \end{cases} \tag{14}$$

However, these thresholding functions have disadvantages that limit their further development. More specifically, the soft thresholding function tends to have a bigger bias, while the hard thresholding function tends to have a bigger variance [45].

There are now many thresholding functions that have been proposed for wavelet applications in signal denoising. Gao et al. [45] proposed a firm thresholding function that remedied the drawbacks of the hard and soft thresholding functions and enabled better denoised results. The firm thresholding function is defined by:

$$T(n) = \begin{cases} 0, & if \ |d_i(n)| \leqslant T_1 \\ sign\{d_i(n)\}\{\frac{T_j(|d_i(n)| - T_1)}{T_j - T_1}\}, & if \ T_1 < |d_i(n)| \leqslant T_j \\ d_i(n), & if \ |d_i(n)| > T_j \end{cases} \tag{15}$$

where T_1 equals $2/3\ T_j$.

The challenge for radar-detected speech is that the noise in radar-detected speech is more complex than traditional microphone speech. Consequently, use of the firm thresholding function results in severe speech distortion. This is because the radar-detected speech consists of predominantly low frequency components, and if all frequency components below a certain threshold are removed, then some essential signal information is also removed.

Yasser [46] proposed a new thresholding function that is given by:

$$T(n) = \begin{cases} d_i(n), & if \ |d_i(n)| \geqslant T_j \\ sign\{d_i(n)\}\{\frac{|d_i(n)|^\gamma}{T_j^{\gamma-1}}\}, & if \ |d_i(n)| < T_j \end{cases} \tag{16}$$

where the γ parameter can be determined by optimization. Although this thresholding function is suitable for radar-detected speech enhancement, it also has a problem in that it preserves any high frequency noise in the detected speech.

In this paper, we propose a modified thresholding function that is intended to solve the listed problems in the two previous functions. The modified function is a combination of the previous two thresholding functions, and is defined as:

$$T(n) = \begin{cases} sign\{d_i(n)\}\{\frac{|d_i(n)|^\gamma}{T_1{}^{\gamma-1}}\}, & if \ \ |d_i(n)| \leqslant T_1 \\ sign\{d_i(n)\}\{\frac{T_j(|d_i(n)|-T_1)}{T_j-T_1}\}, & if\, T_1 < |d_i(n)| \leqslant T_j \\ d_i(n), & if\ |d_i(n)| > T_j \end{cases} \quad (17)$$

The experimental results show that when parameter γ is equal to three, this function provides better results in suppressing the noise radar-detected speech.

4. Results and Discussion

The performance of the proposed algorithm is evaluated in this section. For comparison purposes, two noise suppression algorithms that include spectral subtraction and wavelet shrinkage were also evaluated. Speech time domain waveforms and spectrograms constitute a well-suited tool for evaluating the quality of speech because they can be used to evaluate the extent of the noise reduction, residual noise, and speech distortion by comparing the original radar-detected speech to the enhanced speech. Listening tests were performed to evaluate the performance of the proposed algorithm. In the listening tests, listeners were instructed to evaluate the intelligibility of the original radar-detected speech and the enhanced radar speech based on the criteria of the mean opinion score test (MOS), which is a five-point scale (1: bad; 2: poor; 3: common; 4: good; 5: excellent). All listeners were healthy with no reported history of hearing disorders. In addition, the signal-noise ratio (SNR) was used as an objective measure to evaluate the proposed method's performance. In this section, one sentence of Mandarin Chinese consisting of the words "1-2-3-4-5-6" spoken in sequence was used for evaluation purposes. To guarantee high quality speech signals, a distance of 2 m was selected as the representative distance from loudspeaker to sensor.

4.1. Speech Spectrograms

Figure 5a–d shows the waveforms of the original radar speech, the speech enhanced using the spectral subtraction algorithm, the speech enhanced using the wavelet shrinkage algorithm, and the speech enhanced using the proposed method. Figure 5e–h shows the spectrograms of the original radar speech, the speech enhanced using the spectral subtraction algorithm, the speech enhanced using the wavelet shrinkage algorithm, and the speech enhanced using the proposed method.

Figure 5a,e shows the original radar-detected speech signals. It can be seen from the figures that the original radar-detected speech is contaminated by noise, and the noise is spread across all of the frequency components. As was previously noted, the sources of noise in the radar-detected speech are more complex than those in traditional microphone speech. This is because the sources of noise are a combination of electromagnetic, ambient, and channel noise. Figure 5b,f show the original radar-detected speech after processing using the spectral subtraction method. It can be seen that this algorithm is effective in suppressing the noise in the radar-detected speech both in the speech and non-speech sections. In addition, the noise in all of the frequency components has been removed. However, some new "musical noises" [47], which are similar to the sound of rhythmic music, have been introduced. Consequently, the level of noise reduction achieved using this method is not satisfactory. Figure 5c,g show the original radar-detected speech after processing using the wavelet shrinkage method. It can be seen from the figures that the noises have mostly been removed, especially in the high frequency components. However, there is still too much remnant noise in the low frequency components. Consequently, the quality of the radar-detected speech has not been improved. Figure 5d,h show that the proposed algorithm not only effectively reduces the noise in the radar-detected speech both in the speech and non-speech section, but also reduces the noise across all of the frequency components. These results demonstrate that the proposed method is able to achieve

higher performance than that achieved by spectral subtraction and wavelet shrinkage algorithms because it closely approximates the critical bands of the human auditory system. For these reasons, the proposed method provides an effective way to reduce the noise in radar-detected speech.

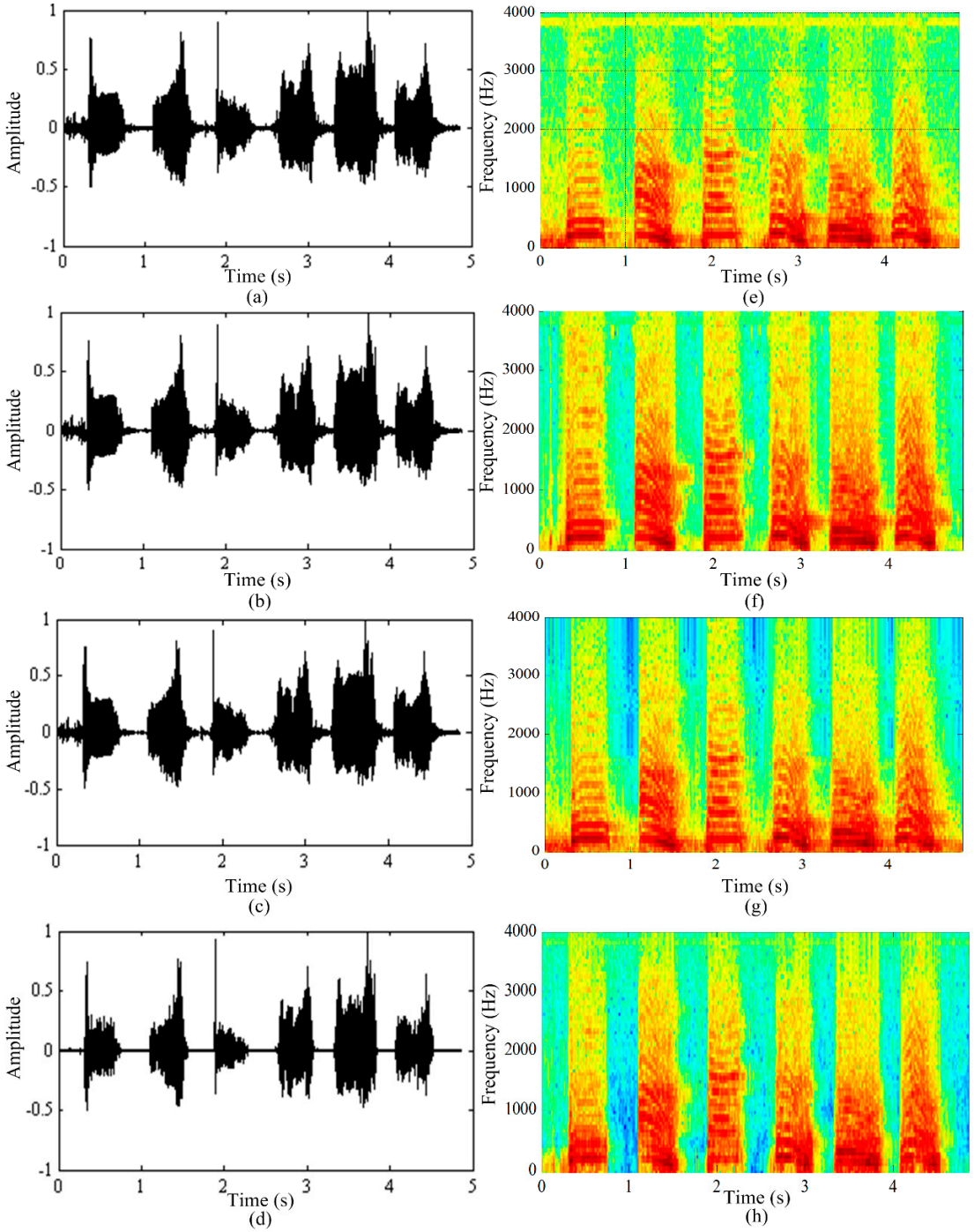

Figure 5. The waveforms and the spectrograms of the radar-detected speech. (**a**,**e**) show the original radar-detected speech; (**b**,**f**) show the enhanced speech obtained using the spectral subtraction method; (**c**,**g**) show the enhanced speech obtained using the wavelet shrinkage method; and (**d**,**h**) show the enhanced speech obtained using the proposed algorithm.

4.2. Listening Tests

In the experiments, three types of noise were selected from the NOISEX-92 database [48], namely white noise, pink noise and babble noise, and added to the original radar-detected speech with SNR inputs of −5, 0, 5, and 10 dB. Four listeners were instructed to evaluate the performance of the three algorithms being tested by listening to the noisy speech signals. The averaged MOS scores for the noisy radar-detected speech and the enhanced noisy radar-detected speech are presented in Figure 6. From the figure, it can be seen that the score of the original noisy speech is approximately "2", and this implies that the quality of the noisy speech is quite poor. However, the score of the enhanced radar speech obtained using the proposed algorithm is acceptable. It also can be seen from the figure that speech signals with higher SNR have correspondingly higher MOS scores For example, for an SNR of 10 dB, the increase in the score for white noise is 0.6; however, the score is only 0.28 when the SNR is −5 dB. This suggests that the proposed algorithm is more suitable in conditions where the SNR is higher. When we compare the three types of noise shown in Figure 6, we find that the MOS scores in the presence of white noise are higher than the scores in the presence of babble and pink noise. This suggests that the proposed algorithm is more "sensitive" to white noise. However, it does not provide satisfactory results in the presence of babble noise.

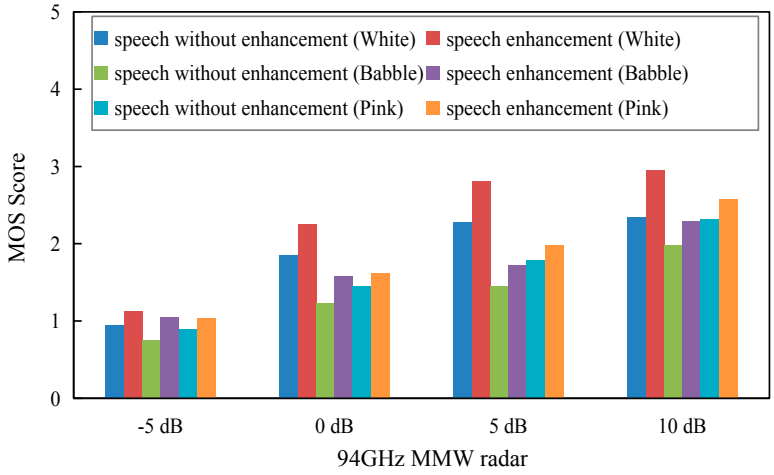

Figure 6. The results of averaged MOS with three types of noise at an SNR of 5 dB.

The radar speech detection experiments demonstrated that the radar-detected speech was shown to have high directional sensitivity and to be immune to strong acoustical disturbance. Therefore, the SNR of the radar-detected speech is expected to be much higher than that of traditional microphone speech. This suggests that the perceptual wavelet packet is suitable for suppressing the noise in radar-detected speech.

The noise in the radar-detected speech was primarily white noise. Thus, in order to further evaluate the performance of the proposed method, ten listeners were selected to listen to the original radar-detected speech and the speech that was enhanced using the three algorithms. The results of the averaged MOS scores are presented in Table 2. The average score for the original radar speech is 3.28. It can be seen from the table that all the scores for the enhanced speech using the three algorithms are improved, and the proposed method achieved the highest score. The MOS score for the spectral subtraction method was lower than both the wavelet shrinkage and the proposed methods. This is because the spectral subtraction method introduced some musical noise into the enhanced speech. The listening tests confirm that the proposed enhancement method is an effective way to suppress the noise in radar-detected speech.

Table 2. Comparison of the MOS obtained by using three enhancement algorithms.

Enhancement Algorithms	Spectral Subtraction	Wavelet Shrinkage	Proposed Method
Averaged MOS	3.71	3.83	3.95
Standard Deviation	0.33	0.27	0.36

4.3. Objective Evaluation

It is possible to determine the amount of noise reduction in a system by measuring the SNR. For this reason, SNR is widely used as an objective method to evaluate the performance of the proposed algorithm. Table 3 shows a comparison between the SNR that was obtained using the different enhancement algorithms on speech signals corrupted by white noise with SNR ranging from −5 to 10 dB. As shown in the Table, the proposed method achieved higher performance than the performance of the other speech enhancement algorithms at the same SNR condition, especially when the incoming SNR was high. The spectral subtraction algorithm achieved a relatively satisfactory result under low SNR conditions; however, the results were poor under high SNR conditions. This is because the algorithm introduced some new "musical noise". The key advantage of the proposed method is that it decomposes the speech signal using a perceptual wavelet packet, which closely approximates the critical bands in the human auditory system. These results suggested that the proposed method is suitable for the enhancement of radar-detected speech.

Table 3. Performance comparison of the SNR obtained for speech signals corrupted by white noise.

Enhancement Algorithms	Noise SNR (dB)				
	0	**5**	**10**	**15**	**20**
Spectral subtraction	6.2	7.1	7.5	7.8	7.9
Wavelet shrinkage	5.4	7.3	10.6	13.3	16.5
Proposed method	6.8	10.7	13.6	15.4	17.1

In order to test the computational complexity of the proposed algorithm, the running time of the three algorithms was computed for the same sentence running on the same hardware resource. The sentence that was chosen was the Mandarin Chinese "1-2-3-4-5-6", and the hardware resource was a Pentium R 3.0 GHz CPU (Intel Corporation, Santa Clara, CA, USA), with 2 GB of RAM (Toshiba Corporation, Minato-ku, Tokyo, Japan). The running time of the spectral subtraction was 0.6428 s, and the running times for the spectral subtraction and wavelet shrinkage were 0.3239 s and 0.3391 s, respectively. These times suggest that the proposed algorithm consumes more hardware resources than the other two algorithms. This is because the proposed method decomposes the speech signal into 17 Bark bands and then removes the noise in each scale. The computational time can be reduced if the decomposition scale is decreased. Based on these results, the proposed algorithm was shown to be an effective way to remove the noise in detected speech signals.

5. Conclusions

This paper proposed a new non-contact speech acquisition and signal processing method that uses a 94 GHz millimeter wave (MMW) radar system and is useful in various applications associated with speech. One problem is that the speech detected using a MMW radar system is often degraded by various sources of noise. Therefore, in order to suppress the noise and enhance the intelligibility of the detected speech, this paper proposed a novel perceptual wavelet packet method.

In the experiments, we found that the proposed algorithm was more "sensitive" to the presence of white noise in the detected speech signals. As the noise in the radar-detected speech is primarily white noise, this new signal processing method is suitable for enhancing the quality of radar-detected speech.

This method is expected to enable promising research and development applications in speech production, speech recognition, and related topics. For example, accurate pitch extraction is always one of the most key issues in speech recognition and synthesis, and the proposed radar system will provide a promising way to extract the pitch directly from the vibration of the vocal cords.

Acknowledgments: This work was supported by the National Natural Science Foundation of China (NSFC, No. 61371163, No. 61327805), the Key Industrial Science and Technology Program of Shaanxi Province, China (No. 2016GY-058), and the High Level Scientific Research Foundation of Xijing University (No. XJ15B01).

Author Contributions: Fuming Chen did the data analysis and prepared the manuscript. Fuming Chen, Sheng Li, Jianqi Wang, Chuantao Li revised and improved the paper. Thanks are given to Qiang An, Fulai Liang and Fugui Qi for the discussion about the method. All authors have read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Scheeper, P.R.; Van der Donk, A.G.H.; Olthuis, W.; Bergveld, P. A review of silicon microphones. *Sens. Actuators A Phys.* **1994**, *44*, 1–11. [CrossRef]
2. Santori, C.M. Bone Conduction Microphone Assembly. U.S. Patent 3,787,641, 22 January 1974.
3. Yanagisawa, T.; Furihata, K. Pickup of speech signal utilization of vibration transducer under high ambient noise. *J. Acoust. Soc. Jpn.* **1975**, *31*, 213–220.
4. Avargel, Y.; Cohen, I. Speech measurements using a laser Doppler vibrometer sensor: Application to speech enhancement. In Proceedings of the Hands-Free Speech Communication and Microphone Arrays (HSCMA), Edinburgh, UK, 30 May–1 June 2011; pp. 109–114.
5. Bakhtiari, S.; Gopalsami, N.; Elmer, T.W.; Raptis, A.C. Millimeter Wave Sensor for Far-Field Standoff Vibrometry. In Proceedings of the 35th Annual Review of Progress in Quantitative Nondestructive Evaluation, Chicago, IL, USA, 20–25 July 2008; Volume 1096, pp. 1641–1648.
6. Li, Z.W. Millimeter wave radar for detecting the speech signal applications. *Int. J. Infrared Millim. Waves* **1996**, *17*, 2175–2183. [CrossRef]
7. Holzrichter, J.F.; Burnett, G.C.; Ng, L.C.; Lea, W.A. Speech articulator measurements using low power EM-wave sensors. *J. Acoust. Soc. Am.* **1998**, *103*, 622–625. [CrossRef] [PubMed]
8. Eid, A.M.; Wallace, J.W. Ultrawideband Speech Sensing. *IEEE Antennas Wirel. Propag. Lett.* **2009**, *8*, 1414–1417. [CrossRef]
9. Lin, C.S.; Chang, S.F.; Chang, C.C.; Lin, C.C. Microwave Human Vocal Vibration Signal Detection Based on Doppler Radar Technology. *IEEE Trans. Microw. Theory Tech.* **2010**, *58*, 2299–2306. [CrossRef]
10. Yueh, S.H.; West, R.; Wilson, W.J.; Li, F.K. Error sources and feasibility for microwave remote sensing of ocean surface salinity. *IEEE Trans. Geosci. Remote Sens.* **2001**, *39*, 1049–1060. [CrossRef]
11. Wang, Y.; Zeng, Y.; Qu, N.; Qu, N.; Zhu, D. Note: Electrochemical etching of cylindrical nanoprobes using a vibrating electrolyte. *Rev. Sci. Instrum.* **2015**, *86*, 076103. [CrossRef] [PubMed]
12. Li, C.; Cummings, J.; Lam, J.; Graves, E.; Wu, W. Radar remote monitoring of vital signs. *IEEE Microw. Mag.* **2009**, *10*, 47–56. [CrossRef]
13. Tian, Y.; Li, S.; Lv, H.; Wang, J.; Jing, X. Smart radar sensor for speech detection and enhancement. *Sens. Actuator A Phys.* **2013**, *191*, 99–104. [CrossRef]
14. Jiao, M.; Lu, G.; Jing, X.; Li, S.; Li, Y.; Wang, J. A novel radar sensor for the non-contact detection of speech signals. *Sensors* **2010**, *10*, 4622–4633. [CrossRef] [PubMed]
15. Obeid, D.; Sadek, S.; Zaharia, G.; El Zein, G. Noncontact heartbeat detection at 2.4, 5.8, and 60 GHz: A comparative study. *Microw. Opt. Technol. Lett.* **2009**, *51*, 666–669. [CrossRef]
16. Mikhelson, I.V.; Bakhtiari, S.; Elmer, T.W., II; Sahakian, A.V. Remote sensing of heart rate and patterns of respiration on a stationary subject using 94-GHz millimeter-wave interferometry. *IEEE Trans. Biomed. Eng.* **2011**, *58*, 1671–1677. [CrossRef] [PubMed]
17. Li, S.; Tian, Y.; Lu, G.; Zhang, Y.; Lv, H.; Yu, X.; Xue, H.; Zhang, H.; Wang, J.; Jing, X. A 94-GHz millimeter-wave sensor for speech signal acquisition. *Sensors* **2013**, *13*, 14248–14260. [CrossRef] [PubMed]
18. Lin, J.C. A new IEEE standard for safety levels with respect to human exposure to radio-frequency radiation. *IEEE Antennas Propag. Mag.* **2006**, *48*, 157–159. [CrossRef]

19. Chen, F.; Li, S.; Li, C.; Liu, M.; Li, Z.; Xue, H.; Jing, X.; Wang, J. A Novel Method for Speech Acquisition and Enhancement by 94 GHz Millimeter-Wave Sensor. *Sensors* **2016**, *16*, 50. [CrossRef] [PubMed]
20. Spriet, A.; Moonen, M.; Wouters, J. Spatially pre-processed speech distortion weighted multi-channel Wiener filtering for noise reduction. *Signal Process.* **2004**, *84*, 2367–2387. [CrossRef]
21. Doclo, S.; Spriet, A.; Wouters, J.; Moonen, M. Speech distortion weighted multichannel Wiener filtering techniques for noise reduction. In *Speech Enhancement*; Springer: Berlin/Heidelberg, Germany, 2005; pp. 199–228.
22. Boll, S.F. Suppression of acoustic noise in speech using spectral subtraction. *IEEE Trans. Acoust. Speech Signal Process.* **1979**, *27*, 113–120. [CrossRef]
23. Goh, Z.; Tan, K.C.; Tan, B.T.G. Postprocessing method for suppressing musical noise generated by spectral subtraction. *IEEE Trans. Speech Audio Process.* **1998**, *6*, 287–292.
24. Li, S.; Wang, J.Q.; Liu, T.; Jing, X.J. Wavelet packet entropy for millimeter wave conducted speech enhancement. *Noise Control Eng. J.* **2009**, *57*, 543–550. [CrossRef]
25. Li, S.; Wang, J.Q.; Jing, X.J. The application of nonlinear spectral subtraction method on millimeter wave conducted speech enhancement. *Math. Probl. Eng.* **2010**, *2010*, 371782. [CrossRef]
26. Kopsinis, Y.; McLaughlin, S. Development of EMD-Based Denoising Methods Inspired by Wavelet Thresholding. *IEEE Trans. Signal Process.* **2009**, *57*, 1351–1362. [CrossRef]
27. Donoho, D.L.; Johnstone, I.M. Adapting to unknown smoothness via wavelet shrinkag. *J. Am. Stat. Assoc.* **1995**, *90*, 1200–1224. [CrossRef]
28. Donoho, D.L.; Johnstone, I.M.; Kerkyacharian, G.; Picard, D. Density estimation by wavelet thresholding. *Ann. Stat.* **1996**, *24*, 508–539. [CrossRef]
29. Donoho, D.L.; Johnstone, I.M. Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **1994**, *81*, 425–455. [CrossRef]
30. Donoho, D.L. De-noising by soft-thresholding. *IEEE Trans. Inf. Theory.* **1995**, *41*, 613–627. [CrossRef]
31. Chambolle, A.; de Vore, R.A.; Lee, N.Y.; Lucier, B.J. Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage. *IEEE Trans. Image Process.* **1998**, *7*, 319–335. [CrossRef] [PubMed]
32. Achim, A.; Tsakalides, P.; Bezerianos, A. SAR image denoising via Bayesian wavelet shrinkage based on heavy-tailed modeling. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 1773–1784. [CrossRef]
33. Bahoura, M.; Rouat, J. Wavelet speech enhancement based on the teager energy operator. *IEEE Signal Process. Lett.* **2001**, *8*, 10–12. [CrossRef]
34. Mercorelli, P. A denoising procedure using wavelet packets for instantaneous detection of pantograph oscillations. *Mech. Syst. Signal Process.* **2013**, *35*, 137–149. [CrossRef]
35. Mercorelli, P. Biorthogonal wavelet trees in the classification of embedded signal classes for intelligent sensors using machine learning applications. *J. Frankl. Inst.* **2007**, *344*, 813–829. [CrossRef]
36. Mercorelli, P. Denoising and harmonic detection using nonorthogonal wavelet packets in industrial applications. *J. Syst. Sci. Complex.* **2007**, *20*, 325–343. [CrossRef]
37. Taşmaz, H.; Erçelebi, E. Speech enhancement based on undecimated wavelet packet-perceptual filterbanks and MMSE–STSA estimation in various noise environments. *Digit. Signal Process.* **2008**, *18*, 797–812. [CrossRef]
38. World Medical Association. World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. *JAMA* **2013**, *310*, 2191–2194.
39. TIMIT Database. Available online: http://www.fon.hum.uva.nl/david/ma_ssp/2007/TIMIT/ (accessed on 15 July 2016).
40. Coifman, R.R.; Meyer, Y. Orthonormal Wave Packet Bases. Preprint, **1990**.
41. Coifman, R.R.; Wickerhauser, M.V. Entropy-based algorithms for best basis selection. *IEEE Trans. Inf. Theory* **1992**, *38*, 713–718. [CrossRef]
42. Chen, S.H.; Wang, J.F. Speech enhancement using perceptual wavelet packet decomposition and teager energy operator. In *Real World Speech Processing*; Springer: New York, NY, USA, 2004; pp. 51–65.
43. Zwicker, E. Subdivision of the audible frequency range into critical bands (Frequenzgruppen). *J. Acoust. Soc. Am.* **1961**, *33*, 248. [CrossRef]
44. Zwicker, E.; Terhardt, E. Analytical Expressions for Critical-Band Rate and Critical Bandwidth as a Function of Frequency. *J. Acoust. Soc. Am.* **1980**, *68*, 1523–1525. [CrossRef]
45. Gao, H.Y.; Bruce, A.G. WaveShrink with firm shrinkage. *Stat. Sin.* **1997**, *7*, 855–874.

46. Ghanbari, Y.; Karami-Mollaei, M.R. A new approach for speech enhancement based on the adaptive thresholding of the wavelet packets. *Speech Commun.* **2006**, *48*, 927–940. [CrossRef]
47. Scalart, P. Speech enhancement based on a priori signal to noise estimation. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Atlanta, GA, USA, 7–10 May 1996; pp. 629–632.
48. NOISEX-92 database. Available online: http://spib.linse.ufsc.br/noise.html (accessed on 15 July 2016).