IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received February 21, 2018, accepted March 21, 2018, date of publication April 2, 2018, date of current version April 23, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2820510

Bloomfield Model Based Signal Process for Networks

CHANGHUA YAO[1], LEI WANG[1], (Member, IEEE), AND XIAOHAN YU[2]

[1]College of Communication Engineering, Army Engineering University of PLA, Nanjing 210007, China
[2]College of Command Information Systems, Army Engineering University of PLA, Nanjing 210007, China

Corresponding author: Lei Wang (iponly@126.com)

This work was supported by the National Natural Science Foundation of China under Grant 61702543 and Grant 71501186.

ABSTRACT This paper proposes a novel speech signal analysis approach based on the Bloomfield (*BF*) model, and provides a formulation of a time-domain *BF* model for speech signals with which speech signals can be reconstructed and the relevant characteristic parameters analyzed. The relationship between the parameters of the *BF* model and those of the linear prediction (*LP*) model are derived, and the speech feature sets derived via the *LP* and *BF* models are compared. A new algorithm is proposed for the recognition of isolated digit speech that utilizes a vector quantization approach and is based on the *BF* Model. The result is obtained with this *BF* approach that provides better results than those of the *LP* model when predicting speech signals. In particular, the *BF* approach has several advantages, including fewer parameters, a lower computational complexity, and accurate characterization of speakers. These advantages ensure the utility of the *BF* model in speech processing applications.

INDEX TERMS Mathematical modeling, bloomfield model, speech signal modeling, speech recognition.

I. INTRODUCTION

The speech signal processing technology is important to the quality of experience in the networks. The efficient speech coding and precise speech conversion is significant to the communication networks limited by the bandwidth and channel quality. The speech signal detection is also critical to the network security in the person identification applications [1], [2]. The method of Linear Prediction (*LP*) analysis is one of the most widely used technique. Many new means and theories have been brought forward since this technique [3]–[6]. The natural basis of the LP model is an all-pole autoregressive moving model *AR*(*p*) [7], [8] when it approximates the original speech signal; however, the *AR* model has some defects that need to be entered in practical applications, such as spectral estimation. There is a big correlation between *AR* spectral resolution and the signal-to-noise ratio (*SNR*) of the input signals, and the order p affects the quality of the spectral estimate. If p increases, the mean square error (*MSE*) of the prediction becomes smaller, accordingly, computational efficiency increases significantly at the same time. The limited capacity of computer processing in networks caused the parameter estimation is also worsened. We can select a more optimum order to minimizing the *MSE* which can achieve a compromise between the computational complexity and precision of the model, the defects of increased computations is still evident [9].

This paper mainly focuses on the technology of speech signal modeling, and can also be used in new areas [10]–[12]. We introduce one new speech analysis approach using Bloomfield's (*BF*) model [13] that addresses the above problems. The simulation results show the *BF* method has a good advantage to curve fitting, spectral estimation, and the modeling of speech signals. We present the results of our experiments on pitch detection and speaker recognition. Even though the *BF* method was proposed in 1973, very little is known about the effect of the use of this method in speech processing in the literature since regarding. Our paper addresses this gap by demonstrating that the *BF* model is broadly applicable in many areas of speech processing.

The remainder of this paper is organized as follows. In Section 2, we review the previous work in Bloomfield's (*BF*) model. Section 3 provides a formulation of a time-domain *BF* model for speech signals. We speech reconstructed signals and analyzed the relevant characteristic parameters. Section 4 proposed a new algorithm based on the *BF* Model for the recognition of isolated digit

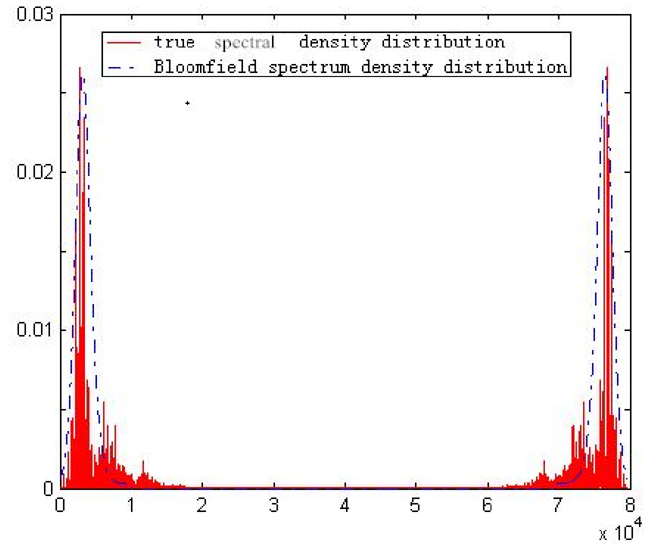


FIGURE 1. Comparison of the actual and BF spectral density distributions.

speech that utilizes a vector quantization (*VQ*) approach, presents experimental show that our approach provides better results than those of the *LP* model. We conclude this paper in Section 5.

II. BLOOMFIELD'S MODEL

The *ARMA* model is well known in linear time sequences. We have found the *BF* model is also convenient.

The idea of estimating the spectrum by Bloomfield in [13] which assumed the time series $\{xt\}$ was stationary. The spectral density function of the series can be written as:

$$S(w) = (2\pi)^{-1}\sigma^2 \exp\left(2\sum_{j=1}^{p}\gamma_j \cos(jw)\right) \quad w \in [-\pi, \pi] \tag{1}$$

where the $\gamma 1, \gamma 2, \cdots, \gamma p, \sigma^2$ are real. The method for determining the order p will be described in detail.

In previous studies, the *BF* model has been ignored by scholars, especially in speech signal processing. It is found in the experiment that the spectral density of the *BF* model is remarkably similar to that of real spectrum density of the speech in Fig. 1. As can be seen from the figure, the two distributions are very similar, which leads us to conclude that the *BF* model is able to accurately describe the spectral density distribution. For this, we have reason to believe the *BF* model can be used in the frequency domain. Analytical methods in the time domain have many advantages, especially since their results can be directly computed. Next, we model signals in the time domain [14].

Suppose $f(z)$ denoted a complex function, it satisfies the following conditions [15]:

(a) $f(0) = 1$;

(b) $f(z) \neq \infty$, $|z| \leq 1$;

(c) $f(z)$ is analytical in the zone $|z| \leq 1, f(e^{i\lambda}) \in L_1$, then

$$f(z) = 1 + f'(0)z + \cdots + \frac{1}{n!}f^{(n)}(0)z^n + \cdots = 1 + \sum_{n=1}^{\infty}(1/n!)f^{(n)}(0)z^n \tag{2}$$

where the variable z is a backward-shift operator, which can be defined as follows:

$$f(z)X_t = X_t + \sum_{i=1}^{\infty}(1/i!)f^{(i)}(0)X_{t-i} \tag{3}$$

where X_t is an arbitrary stationary time series.

If we assume:

(d) $(1/n!)f^{(n)}(0) = o(n^{-s}),\ s > 1/2$;

Then the above series converges in the sense of the mean square, i.e., there exists a stationary time series Y_t that allows:

$$E\left|Y_t - X_t - \sum_{i=1}^{N}\frac{1}{i!}f^{(i)}(0)X_{t-i}\right|^2 \rightarrow 0, \quad N \rightarrow \infty.$$

Thus, we can denote

$$f(z)X_t = Y_t \tag{4}$$

If we assume that:

(e). $(1/n!)f^{(n)}(0) = o(n^{-s}), \quad s > 1$.

Then the above infinite series converges almost everywhere on almost every track. If we choose limits for Yt, the same relationship can be obtained as in Eq. (4).

In particular, if Xt is a white noise sequence, then Yt is normally a physically realizable generic linear sequence that satisfies conditions (a), (b), (c), and (d), or a linear sequence that satisfies conditions (a), (b), (c), (d), and (e). The latter case also allows $f^{-1}(z)Yt = Xt$. On the other hand, if we let εt represent white noise, and

$$X_t = \sum_{j=1}^{\infty}a_j\varepsilon_{t-j}$$

be a physically realizable linear sequence with $a0 = 1$ and $\sum|aj| < \infty$, then series then series$\sum a_jz^j$ uniformly converges in $|z| \leq 1$, which is denoted as $g(z)$.

Thus, according to the above convention, X_t can be expressed as $X_t = g(z)\varepsilon_t$. Furthermore, if we assume $\sum a_jz^j \neq 0$ with $|z| < 1$, we obtain $\varepsilon_t = g^{-1}(z)X_t$.

It can be shown that, on the one hand, physically realizable linear sequences Xt can be represented in the form $Xt = f(z)\varepsilon t$, while on the other hand, the time series of Eq. (4) is a physically realizable linear stationary time series if a certain condition $f(z)$ is satisfied. So, Eq. (4) can be referred to as the generic linear time series model, and the research on Xt can be substituted with that of $f(z)$.

The equivalent of the model is $(f(z))^{-1}Xt = \varepsilon t$, and model (5) in [13], is continuous, and the spectral density can be written:

$$S(w) = (2\pi)^{-1}\sigma^2\left|f(e^{-iw})\right|^2 \tag{5}$$

Suppose $f(z)$ is a polynomial and it is rootless, Eq. (4) is considered to be $MA(p)$, and $f(z)Xt = \varepsilon t$ is considered to be $AR(p)$. Model (4) can be considered to be an ARMA.

Let us consider Eq. (4). Suppose $f(z)$ satisfies assumptions (a), (b), (c), and (d), ε_t is the same as previous, and $G_n = (1/n!)f^{(n)}(0)$, then $\{G_n\}$ has the same meaning as the usual Green function. Therefore, we refer to it as the Green function of Eq. (3). The correlation function of X_t can be calculated as follows:

$$\begin{aligned}\gamma(\tau) &= E(X_t X_{t+\tau}) \\ &= E[(\sum_{k=0}^{\infty} G_k \varepsilon_{t-k})\overline{(\sum_{j=0}^{\infty} G_j \varepsilon_{t-j+\tau})}] \\ &= \sigma^2 \sum_{k,j} G_k \overline{G_k} \delta(-k, \tau - j) \\ &= \sigma^2 \sum_{k=0}^{\infty} G_k \overline{G_{k+\tau}}\end{aligned}$$

where δ satisfies:

$$\delta(i,j) = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases}$$

In particular, if $f(z)$ is located on the real axis, then we have:

$$\gamma(\tau) = \sigma^2 \sum_{k=0}^{\infty} G_k G_{k+\tau}, \quad \gamma(0) = \sigma^2 \sum_{k=0}^{\infty} G_k^2 \tag{6}$$

On the surface, for a model with the general form of (6), its characteristics are not able to be studied in detail. However, with limited parameters and equations in known forms, model (6) can still be used in many cases to solve practical problems.

In the next section, we establish the time domain *BF* model.If the *BF* model satisfies conditions (a), (b), and (c), the time sequence Xt determined by

$$X_t = \exp\left(\sum_{j=1}^{p} \gamma_j z^j\right) \varepsilon_t$$

can be considered to be a stable linear zero-mean time sequence, so the time-domain *BF* model is as follows:

$$X_t = \exp\left(\sum_{j=1}^{p} \gamma_j z^j\right) \varepsilon_t.$$

However, the equivalent model in Eq. (7) is sometimes used instead.

$$\exp\left(-\sum_{j=1}^{p} \gamma_j z^j\right) X_t = \varepsilon_t \tag{7}$$

III. DERIVING SPEECH PARMATERS BASAED ON THE BF MODE

A. LINEAR PREDICTIVE ANALYSIS OF A SPEECH SIGNAL

In modeling a stationary time sequence using the *BF* model, we estimate $\gamma 1, \gamma 2, \cdots, \gamma p, \sigma^2$ [15] using $x1, x2, \ldots .xn$. When p has been selected, the estimation formula of γj and σ^2 are as follows:

$$\hat{\gamma_j} = \frac{2}{N} \sum_{k=1}^{N_0} [\log I_N(2kN^{-1}\pi)] \cos(2kN^{-1}j\pi) \tag{8}$$

$$\hat{\sigma}^2 = 2\pi \exp\left(0.57722 + 2N^{-1} \sum_{k=1}^{N_0} \log I_N(2kN^{-1}\pi) \right) \tag{9}$$

where $N0 = ((N-1)/2)$, 0.55772 called *Euler's* constant, and the period gram is:

$$I_N(\lambda) = \frac{1}{2\pi N} \left| \sum_{k=1}^{N} X_k \exp(-ik\lambda) \right|^2, \quad \lambda \in [-\pi, \pi] \tag{10}$$

we use the following formula to calculate the order p:

$$AIC(s) = 2\pi \hat{\sigma}^{-2} \int_{-\pi}^{\pi} I_N(\lambda) \exp\left(-2 \sum_{j==1}^{s} \hat{\gamma_j} \cos j\lambda \right) d\lambda + 2s/N \tag{11}$$

In a speech signal, there is a large correlation between the adjacent sample values. The signal value at a certain time can be predicted based on the past sampled values, i.e., each sample value can be approximated by linear combinations of several past sampled values:

$$s(n) = Ge(n) + \sum_{i=1}^{p} a_i s(n-i) \tag{12}$$

The purpose of linear predictive analysis is to derive the prediction coefficients under the minimum mean-square error criterion. The error $\varepsilon(n)$ is:

$$\varepsilon(n) = s(n) - \sum_{i=1}^{p} a_i s(n-i) \tag{13}$$

B. DERIVING THE LP PARAMETER

Speech signals can be described by various parameters. By using the *LP* method to analyze the speech signal in a frame, a set of *LP* parameters and features can be derived that are applicable to different aspects of speech signal processing. These features are the Linear Predictive Cepstral Coefficients (*LPCCs*), part of the correlation coefficient or reflection coefficient (*REFL*) ($ki, i = 1, 2, \cdots, p$), the log area ratio (*LAR*) coefficient ($gi, i = 1, 2, \cdots, p$), the coefficient of sine (*ARCSIN*), the line spectral frequencies (LSF), etc. *BF* model estimation is based on speech sampling point data, which utilizes parameters, and we can derive the equivalent *LP* model ($\{a1, a2, \ldots ap\}$) and other parameters from the model features (γ_i, σ^2).

We now compare the time domain *BF*-model in Eq. (7) and $\{a1, a2, \ldots ap\}$ derived from Eq. (13) with the mean square error criterion. In both the time domain BF model and *LP* models, the current sample value is related to the past samples. However, the *LP* model is linear while the *BF* model is exponential. In the *LP* model, the linear prediction coefficients $\{a1, a2, \ldots ap\}$ are derived via the smallest mean square criterion. Nevertheless, the predicted coefficients in the *BF* model are obtained by calculating the parameters γj when the model is adapted to a form similar to that of the *LP* model, which means that the prediction coefficients are derived from γj and the inference criterion is a criterion of the model itself. Because speech signals are correlated over short periods of time, the predicted coefficients obtained from the two models converge within the allowable range of error, as long as the time scope and order are reasonable.

We now approximate the 10th-order *LP* model using a second-order Taylor expansion of the *BF* model, and obtain the following:

$$\begin{aligned}
a_1 &= \gamma_1, \\
a_2 &= -\frac{\gamma_1^2}{2} + \gamma_2, \\
a_3 &= \frac{\gamma_1^3}{6} - \gamma_1\gamma_2, \\
a_4 &= -\frac{\gamma_1^4}{24} + \frac{\gamma_1^2\gamma_2}{2} - \frac{\gamma_2^2}{2}, \\
a_5 &= \frac{\gamma_1^5}{120} - \frac{\gamma_1^3\gamma_2}{6} + \frac{\gamma_1\gamma_2^2}{2}, \\
a_6 &= -\frac{\gamma_1^6}{720} + \frac{\gamma_1^4\gamma_2}{24} - \frac{\gamma_1^2\gamma_2^2}{4} + \frac{\gamma_2^3}{6}, \\
a_7 &= \frac{\gamma_1^7}{5040} - \frac{\gamma_1^5\gamma_2}{120} + \frac{\gamma_1^3\gamma_2^2}{12} - \frac{\gamma_1\gamma_2^3}{6}, \\
a_8 &= -\frac{\gamma_1^8}{40320} + \frac{\gamma_1^6\gamma_2}{720} - \frac{\gamma_1^4\gamma_2^2}{48} + \frac{\gamma_1^2\gamma_2^3}{12} - \frac{\gamma_2^4}{24}, \\
a_9 &= \frac{\gamma_1^9}{362880} - \frac{\gamma_1^7\gamma_2}{5040} + \frac{\gamma_1^5\gamma_2^2}{240} - \frac{\gamma_1^3\gamma_2^3}{36} + \frac{\gamma_1\gamma_2^4}{24}, \\
a_{10} &= -\frac{\gamma_1^{10}}{3628800} + \frac{\gamma_1^8\gamma_2}{40320} + \frac{\gamma_1^6\gamma_2^2}{1440} + \frac{\gamma_1^4\gamma_2^3}{144} \\
&\quad - \frac{\gamma_1^2\gamma_2^4}{48} + \frac{\gamma_2^5}{120} \qquad (14)
\end{aligned}$$

In Eq. (14), the 10 parameters of the *LP* model are represented by two parameters of the *BF* model, which greatly reduces the computational complexity. In the next section, we evaluate new model and illustrate applications in speech signal processing.

IV. APPLICABILITY OF THE BF MODEL TO SPEECH PROCESSING

A. COMPARISON OF SPEECH PREDICTION WAVEFORMS

Many new techniques have been used In parametric speech synthesis over the past decade [17]–[20]. The literature [21] proposed a time-domain deterministic plus noise model based hybrid source modeling framework. In this paper, we model the original speech signal using time-domain *BF* model. In this experiment, A voice segment is sampled. We illustrates results of the $3-order$ *BF* model and $12-order$ *LP* model. The experiment shows that the waveform fitting with *BF* model is closer to the original voice waveform in the peaks and valleys. *BF* model has a good advantage with fewer parameters and a lower computational complexity.

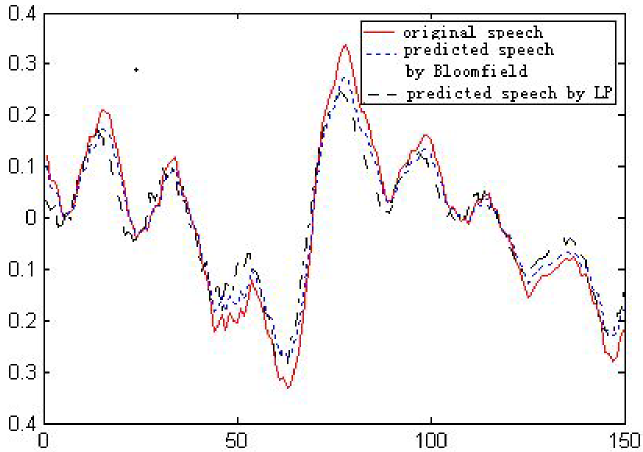


FIGURE 2. Speech waveforms comparison.

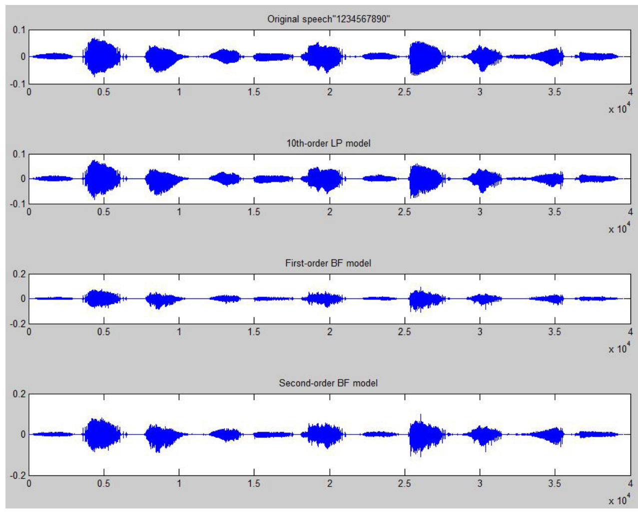
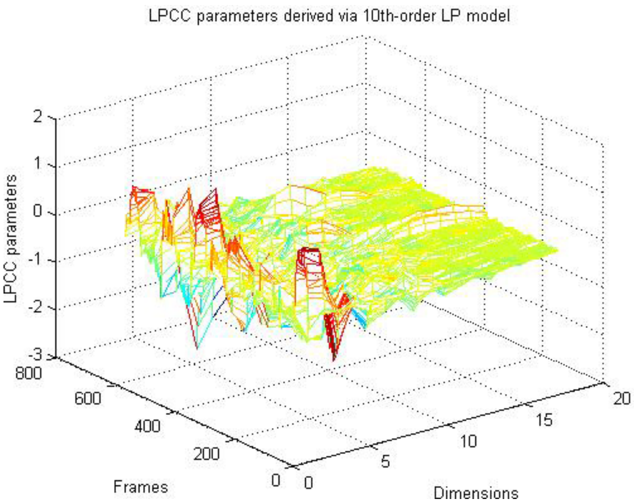


FIGURE 3. Speech signal reconstruction via a 10th-order LP model, and first- and second-order BF models.

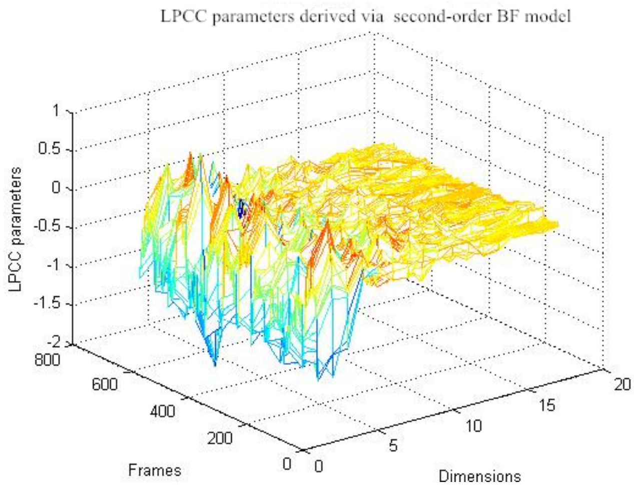
The speech samples are of a female voice reciting "1234567890," and a sampling frequency of 8 KHz was used. The following processing steps were performed on the input speech: pre-emphasis using a $1 - 0.975z^{-1}$ filter, windowing, endpoint detection, setting the frame length to 128 points, moving the frame through 64 points, and then reconstructing the voice signal via frame-by-frame calculation using the *BF* and *LP* models. Shown in Fig. 3 are the "1234567890" speech signal reconstructions using a 10th-order *LP* model, and first- and second-order *BF* models, respectively. The results show that, compared to the traditional *LP* model, the *BF* model restores the speech with very little cost, i.e., little distortion.

B. SPEECH SIGNAL FEATURE EXTRACTION

Ooi Chia Ai discussed the comparison of speech parameterization methods: mel-frequency cepstrum coefficients $(MFCC)$ and linear prediction cepstrum coefficients $(LPCC)$

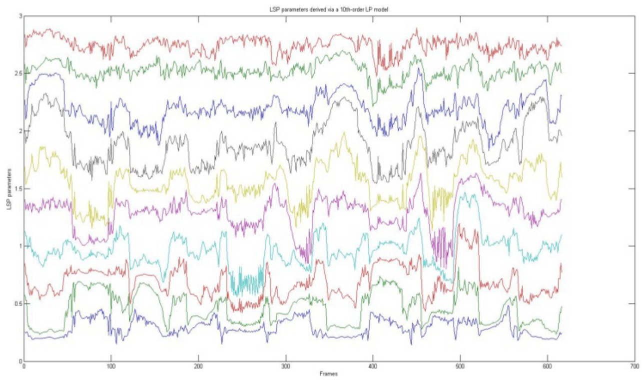


(a)

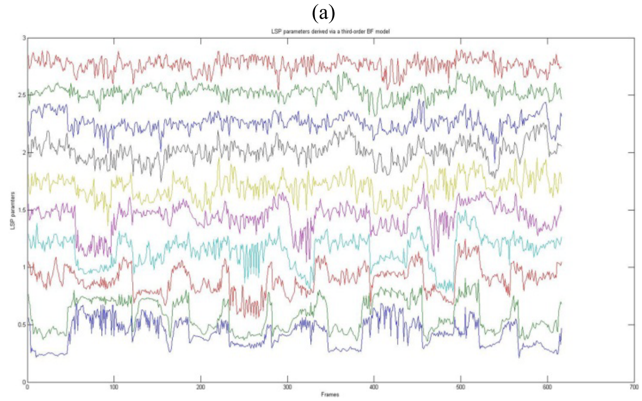


(b)

FIGURE 4. (a) MLPCC parameters derived from a 10th-order LP model. (b) LPCC parameters derived from a second-order BF model.



(a)



(b)

FIGURE 5. (a) LSP parameters derived 10th-order LP model. (b) LSP parameters derived third-order BF model.

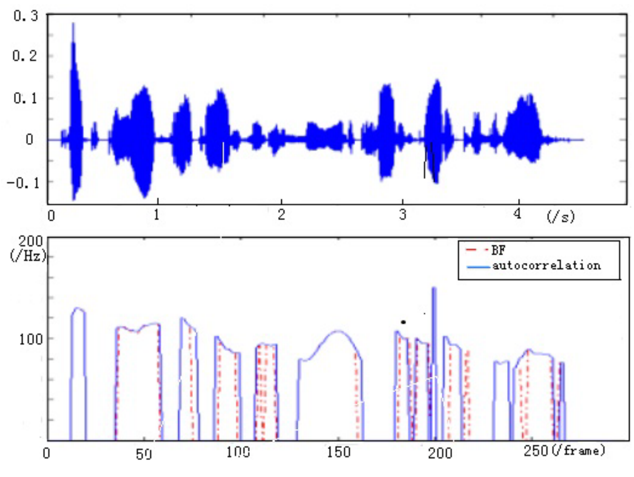


FIGURE 6. The the comparison diagram of pitch trajectories using Two models.

for recognizing stuttered events. The experimental investigation determined that the *LPCC* features slightly outperformed the *MFCC* features [22]. In the *LP* model, other equivalent speech parameters can be derived. They are the *LPCC*, part of the correlation coefficient or reflectance (*REFL*) ($ki, i = 1, 2, \ldots, p$), the log area ratio coefficient (*LAR*) ($gi, i = 1, 2, \ldots, p$), the coefficient of sine (*ARCSIN*), and the line spectrum pair (*LSP*). According to the algorithm of this paper, the 18th-order LPCC parameters can be derived, which are shown in Fig. 4, where the x-axis represents the cepstrum coefficient, the y-axis represents an experimental voice frame, the z-axis represents the corresponding value of the cepstral, and the *LPCC* parameter is 17 ($c1, \cdots c18$). Considering the case where the value of the first-dimension energy c0 is very large. Generally, in speech recognition systems, c0 is called the energy factor and not the cepstrum. In the second-order *BF* model reconstructed speech, there are more LPCC burrs, and the noise immunity is worse. This is because the error is large when the *LP* model is approximated with a small order *BF* model.

Fig. 5 shows the LSP parameters derived from a 10-order *LP* model, and first- and third-order *BF* models, respectively. Our experiments show that the LSP parameters derived from the *BF* model exhibited more detailed characteristics and better frequency resolution than those from the *LP* model [23].

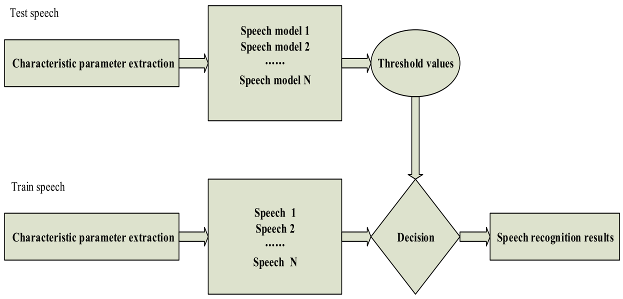


FIGURE 7. Speech recognition system based on the VQ algorithm.

Furthermore, the frequency resolution increased when the order of the model increased. Nevertheless, the reflected channel amplitude spectrum on the *LP* model is better than that of the *BF* model, i.e., the *LP* model better represents the overall frequency trend.

C. ESTIMATE THE SPEECH PICH USING THE BF MODEL

Until now, the Voiced/Unvoiced determination is an important issue in pitch detection. An improved method is put forwarded to estimate the speech pitch, where the speech would be inverse-filtered before the pitch is observed. The simulation results of the new method are shown to be smoother. The general form of inverse filter:

$$A(z) = 1 + \sum_{i=1}^{\overline{M}} a_i z^{-i}$$

is defined by determining $\{ai\}$, that is achieved by *BF* model. The model would produce the time-domain waveform. As a result, it is alike for the two-pitch trajectory from [24] (see Fig. 6).

In [21], the error detection rate (*EDR*) and valid pitch err (*VPE*) would be adopted to show the performances of the traditional autocorrelation and the improved

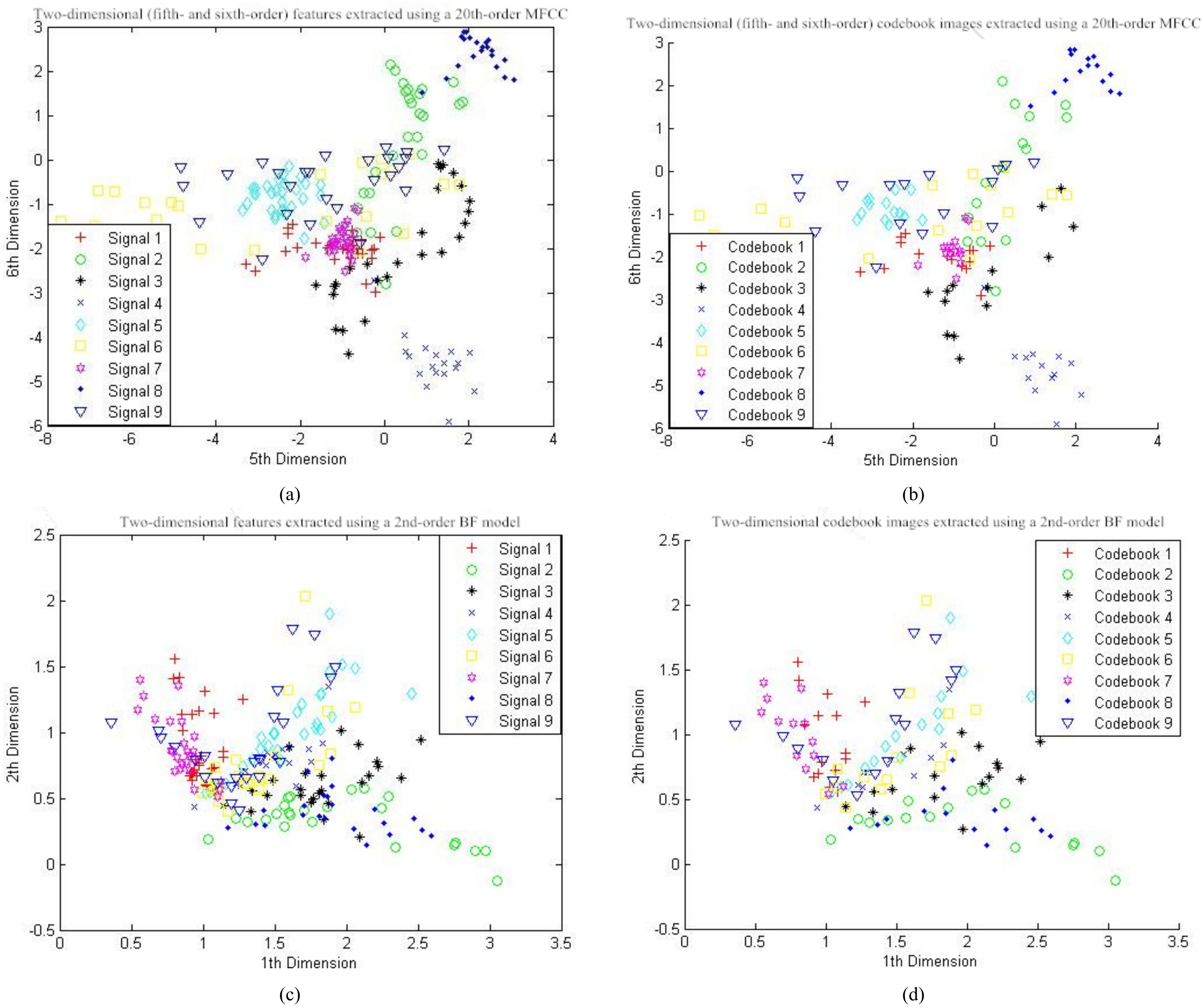


FIGURE 8. Two-dimensional features and codebook images from a 20th-order MFCC and second-order BF model. (a) Features from the 20th-order MFCC. (b) Codebook images from the 20th-order MFCC. (c) MFCC Features from the second-order BF model. (d) Codebook images from the second-order BF.

TABLE 1. The comparison of two model.

	BF model	Autocorrelation model
$EDR(\%)$	2.47	3.80
$VPE(/MS)$	0.039	0.035

approach. Table 1 shows, as the improved approach is worse than that the traditional approach on VPE, but the improved approach has better EDR performance, and the new model generally performs better. This also verifies the positive determination of the previous work [24].

D. SPEECH IDENTIFICATION USING THE BF MOEDL

Speech recognition technology is an important part of speech signal processing, and is widely used as an effective interface in many devices, including personal computers, robots, mobile phones, and vehicle navigation systems [25]–[28]. An important aspect of speech recognition is feature parameter extraction, i.e., identifying a group of voice features that are robust and less prone to interference, and can maintain a certain level of recognition performance with different speakers and background noise. Our experiments began with voice signal pretreatment, extraction of the mel-cepstrum coefficients and *BF* parameters of speech, and isolated-word recognition. Then, we compared the recognition results to verify that the *BF* model parameters are applicable to speech recognition.

As for the issues in speech identification, because the purpose of this paper is to validate the feasibility of the parameters of the *BF* model, we use small vocabularies and an isolated-Word vector quantization (VQ) method to implement the speech recognition simulation. Fig. 7 shows a diagram of the speech recognition system based on the VQ algorithm.

In order to obtain experimental data, four college students (two boys and two girls, each approximately 20 years of age) participated by recording their voices and setting up a voice library. The voice-recording software used was Cool Edit Pro 2.0. All the recorded data used a sampling rate of 8 KHz, a mono voice channel, 16-bit quantization, and was stored on a PC machine in wave (*.wav) format. The voice sample base used in the experiment was based on Chinese digits (1–9), pronounced 270 times per person (each digit was pronounced 30 times), for a total of 1,080 samples. Of these, 540 samples were used as the training samples, and the other half were used as the testing samples. After preprocessing, feature extraction on the training voice samples was performed with 256 samples per frame, followed by a frame shift of 100, before extracting the coefficients of the 20th-order MFCC [29] and the parameters of the second-order BF.

Because the *BF* model is of order two, i.e., it extracts two phonological parameters per frame, two dimensions of MFCC coefficients are also extracted for comparison purposes. Over the course of the trials, it was found that the two-dimensional image constructed from the fifth and sixth dimensions of the 20th-order MFCC coefficients reasonably reflects the characteristics of the speech signals. To use the VQ method in the *BF* model, the Linde–Buzo–Gray (LBG) algorithm was utilized with the two-dimensional values clustered from the same fifth- and sixth-dimensional coefficients to characterize the feature parameters of the speech. In Fig. 8, where the fifth and sixth dimensions of the 20th-order MFCC and second-order *BF* model parameters are used as coordinates, the two-dimensional images depicted by the codebooks clustered using the LBG algorithm are shown.

V. CONCLUSION

In this paper, we systematically applied Bloomfield's model to speech processing and formulated a model of the speech signal. We introduced the basic concept and theory of the *BF* model, reconstructed a speech signal based on the *BF* model, and analyzed the relevant characteristic parameters. We derived the relationship between the parameters of the *BF* and *LP* models, and surveyed the *LP*- and *BF*-derived feature sets. We proposed a new algorithm for isolated digit speech recognition based on the *BF* Model by utilizing easily realized VQ techniques, and validated the applicability of the new model for speech recognition. The experiments show that the *BF* model of speech signals is useful and has excellent characteristics, such as being able to accurately synthesize voice signals and perform small vocabulary isolated-word recognition tasks with only a few parameters. The experimental results of the *BF* model compared favorably to those of the *LP* model. Based on these results, this model should prove useful in a variety of applications; however, further development is required. In the future, we intend to study the applicability of the model to efficient speech coding schemes and precise speech conversion techniques.

REFERENCES

[1] W. Wei *et al.*, "Gradient-driven parking navigation using a continuous information potential field based on wireless sensor network," *Inf. Sci.*, vol. 408, pp. 100–114, Oct. 2017.

[2] Y. G. Sun, H. Y. Qiang, X. Mei, and Y. Teng, "Modified repetitive learning control with unidirectional control input for uncertain nonlinear systems," *Neural Comput. Appl.*, pp. 1–10, Apr. 2017, doi: 10.1007/s00521-017-2983-y.

[3] R. Peng, Z.-H. Tan, X. Li, and C. Zheng, "A perceptually motivated LP residual estimator in noisy and reverberant environments," *Speech Commun.*, vol. 96, pp. 129–141, Feb. 2018.

[4] T. L. Jensen, D. Giacobello, T. van Waterschoot, and M. G. Christensen, "Fast algorithms for high-order sparse linear prediction with applications to speech processing," *Speech Commun.*, vol. 76, pp. 143–156, Feb. 2016.

[5] N. Nower, Y. Liu, and M. Unoki, "Restoration scheme of instantaneous amplitude and phase using Kalman filter with efficient linear prediction for speech enhancement," *Speech Commun.*, vol. 70, pp. 13–27, Jun. 2015.

[6] M. Parchami, W.-P. Zhu, and B. Champagne, "Speech dereverberation using weighted prediction error with correlated inter-frame speech components," *Speech Commun.*, vol. 87, pp. 49–57, Mar. 2017.

[7] Y. Shekofteh and F. Almasganj, "Autoregressive modeling of speech trajectory transformed to the reconstructed phase space for ASR purposes," *Digit. Signal Process.*, vol. 23, no. 6, pp. 1923–1932, 2013.

[8] L. Legrand and E. Grivel, "Jeffrey's divergence between moving-average models that are real or complex, noise-free or disturbed by additive white noises," *Signal Process.*, vol. 131, pp. 350–363, Feb. 2017.

[9] M. Nasrolahzadeh, Z. Mohammadpoori, and J. Haddadnia, "Analysis of mean square error surface and its corresponding contour plots of spontaneous speech signals in Alzheimer's disease with adaptive Wiener filter," *Comput. Human Behavior*, vol. 61, pp. 364–371, Aug. 2016.

[10] A. Yang *et al.*, "Optimum surface roughness prediction for titanium alloy by adopting response surface methodology," *Results Phys.*, vol. 7, pp. 1046–1050, 2017.

[11] K. Cui, W.-H. Yang, and H.-Y. Gou, "Experimental research and finite element analysis on the dynamic characteristics of concrete steel bridges with multi-cracks," *J. Vibroeng.*, vol. 19, no. 6, pp. 4198–4209, 2017.

[12] K. Cui and X. Qin, "Virtual reality research of the dynamic characteristics of soft soil under metro vibration loads based on BP neural networks," *Neural Comput. Appl.*, vol. 29, no. 5, pp. 1233–1242, 2018, doi: 10.1007/s00521-017-2853-7.

[13] P. Bloomfield, "An exponential model for the spectrum of a scalar time series," *Biolmetrika*, vol. 60, no. 2, pp. 217–226, 1973.

[14] Z. J. Xie, *The Time Series Analysis*. Beijing, China: Beijing University, 1990, pp. 223–230.

[15] Z. Yue, Z. Cao, and H. Chen, "A linear time sequence model with infinitude order," *J. Southeast Univ.*, vol. 24, no. 2, pp. 54–59, Mar. 1994.

[16] G. Wng and Y. Zhong, "Some new progress of support vector machine," *Chin. J. Electron.*, vol. 20, no. 10, pp. 233–243, Feb. 2001.

[17] K. Tokuda, Y. Nankaku, T. Toda, H. Zen, J. Yamagishi, and K. Oura, "Speech synthesis based on hidden Markov models," *Proc. IEEE*, vol. 101, no. 5, pp. 1234–1252, May 2013.

[18] Y. Tang, M. Cooke, and C. Valentini-Botinhao, "Evaluating the predictions of objective intelligibility metrics for modified and synthetic speech," *Comput. Speech Lang.*, vol. 35, pp. 73–92, Jan. 2015.

[19] T. Nagata, H. Mori, and T. Nose, "Dimensional paralinguistic information control based on multiple-regression HSMM for spontaneous dialogue speech synthesis with robust parameter estimation," *Speech Commun.*, vol. 88, pp. 137–148, Apr. 2017.

[20] T. Raitio, L. Juvela, A. Suni, M. Vainio, and P. Alku, "Phase perception of the glottal excitation and its relevance in statistical parametric speech synthesis," *Speech Commun.*, vol. 81, pp. 104–119, Jul. 2016.

[21] N. P. Narendra and K. S. Rao, "Time-domain deterministic plus noise model based hybrid source modeling for statistical parametric speech synthesis," *Speech Commun.*, vol. 77, pp. 65–83, Mar. 2016.

[22] O. C. Ai, M. Hariharan, S. Yaacob, and L. S. Chee, "Classification of speech dysfluencies with MFCC and LPCC features," *Expert Syst. Appl.*, vol. 39, no. 2, pp. 2157–2165, 2012.

[23] X.-Q. Zhao, Y.-N. Liu, and R. Wu, "A new approach of pitch detection based on morphology filter," *J. China Inst. Commun.*, vol. 25, no. 5, pp. 53–58, 2004.

[24] Z. J. Yue, L. Wang, and H. Wang, "A new parameter of speech character based on the Bloomfield's model," in *Proc. CISP*, Tianjin, China, Oct. 2009, pp. 1–4.

[25] A. M. C. Martinez, S. H. Mallidi, and B. T. Meyer, "On the relevance of auditory-based Gabor features for deep learning in robust speech recognition," *Comput. Speech Lang.*, vol. 45, pp. 21–38, Sep. 2017.

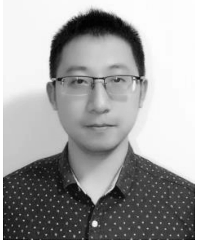
[26] T. Grozdić, S. T. Jovičić, and M. Subotić, "Whispered speech recognition using deep denoising autoencoder," *Eng. Appl. Artif. Intell.*, vol. 59, pp. 15–22, Mar. 2017.

[27] A. E. A. Ferreira and D. Alarcão, "Real-time blind source separation system with applications to distant speech recognition," *Appl. Acoust.*, vol. 113, pp. 170–184, Dec. 2016.

[28] Y.-G. Sun, H.-Y. Qiang, J.-Q. Xu, and D.-S. Dong, "The nonlinear dynamics and anti-sway tracking control for offshore container crane on a mobile harbor," *J. Marine Sci. Technol.*, vol. 25, no. 6, pp. 656–665, 2012.

[29] A. O. Aliyu, E. O. Adewale, and O. S. Adetunmbi, "Development of a text-dependent speaker recognition system," *Int. J. Comput. Appl.*, vol. 69, no. 16, pp. 1–7, 2013.

CHANGHUA YAO received the B.S. degree in automation from Zhejiang University in 2005 and the Ph.D. degree in communications and information systems from the PLA University of Science and Technology in 2016. He has been holding a Postdoctoral position with the Army Engineering University of PLA since 2017. He has authored over 25 scientific articles. His research interests focus on wireless networks, network security, opportunistic spectrum access, distributed optimization, data analysis, artificial intelligence, and machine learning.



LEI WANG received the Ph.D. degree in military operational research from the PLA University of Science and Technology, in 2014, China. He currently holds a Postdoctoral position of communication and information system, an Engineer of the optimization and system engineering with the College of Communications Engineering, Army Engineering University of PLA, China. He has authored over 25 scientific articles. His research interests focus on knowledge engineering, data mining, artificial intelligence, network planning, and military operational research.

XIAOHAN YU received the Ph.D. degree from the College of Communications Engineering, PLA University of Science and Technology, in 2014. He is currently a Lecturer with the College of Command and Control Systems, Army Engineering University of PLA. He has authored over 30 scientific articles. His research interests focus on multicriteria decision making, data mining, and military operational research.