

# Chapter 54

## Throat Polyps Detection Based on Patient Voices

Zhen Zhong, Zhangliang Chen, Qilian Liang, and Shuifang Xiao

**Abstract** In this paper, we present a new approach for throat polyps detection based on patient's vowel voices using fuzzy classifiers. Based on human voice samples and Hidden Markov Model, we show that transformed voice samples (linearly combined samples) follow Gaussian distribution, further we demonstrate that a type-2 fuzzy membership function (MF), i.e., a Gaussian MF with uncertain mean, is most appropriate to model the transformed voices samples. We also apply Short-Time-Fourier-Transform (STFT) and Singular-Value-Decomposition (SVD) to the vowel voice samples, and observe that the power decay rate could be used as an identifier in throat polyps detection. Two fuzzy classifiers and a Bayesian classifier are designed for throat polyps detection based on human vowel voices /a:/ and /i:/ only, and the fuzzy classifiers are compared against the Bayesian classifier. Simulation results show that an interval type-2 fuzzy classifier performs the best of the three classifiers.

**Keywords** Polyps detection • Fuzzy logic systems • Bayesian classifier • Interval type-2 fuzzy classifier • Fuzzy membership functions

---

Z. Zhong (✉)

Dept of ENT and Neck Surgery, Peking University First Hospital, Beijing 100034, China  
e-mail: [zhong\\_zhen@sina.com](mailto:zhong_zhen@sina.com)

Z. Chen

College of Precision Instrument and Opto-Electronics Engineering, Tianjin University, Tianjin 300072, China  
e-mail: [tjuczl@126.com](mailto:tjuczl@126.com)

Q. Liang

Dept of Electrical Engineering, University of Texas at Arlington, Arlington, TX 76019, USA  
e-mail: [liang@uta.edu](mailto:liang@uta.edu)

S. Xiao

Dept of ENT and Neck Surgery, Peking University First Hospital, Beijing 100034, China  
e-mail: [xiao\\_ent@yahoo.com.cn](mailto:xiao_ent@yahoo.com.cn)

## 54.1 Introduction

The throat polyps detection is a field which demands more investigation. Traditionally, the methods of diagnosis are indirect laryngoscope, video-laryngoscope, and stroboscope light [1]. However, most of these methods need special instrument, and mainly depend on the experience of the pathologists. It would be desirable if throat polyps could be detected based on the patient vowel voices only. Traditional pattern recognition techniques such as Bayesian classifier, known as the optimal classifier, could be used if the voice samples follow certain distribution, and this belongs to model-based statistical processing. In human's voices, the voice amplitude is highly bursty, and we believe that no statistical model can really demonstrate the uncertain nature of the voice. Fuzzy logic systems (FLS) are model free. Their membership functions are not based on statistical distributions. In this paper, we, therefore, apply fuzzy techniques to polyps patient diagnosis.

In Sect. 54.2, we model voice samples using interval type-2 Gaussian membership function. In Sect. 54.3, we apply STFT and SVD to voice samples. In Sect. 54.4, a Bayesian classifier is proposed. Performances of the three classifiers are evaluated in Sect. 54.5. Conclusions are presented in Sect. 54.6.

## 54.2 Modeling Voice Samples Using Hidden Markov Model and Gaussian Primary MF with Uncertain Mean

In [3], autoregressive Hidden Markov Model (HMM) was used to represent voice samples  $x_i$ , which means we could have

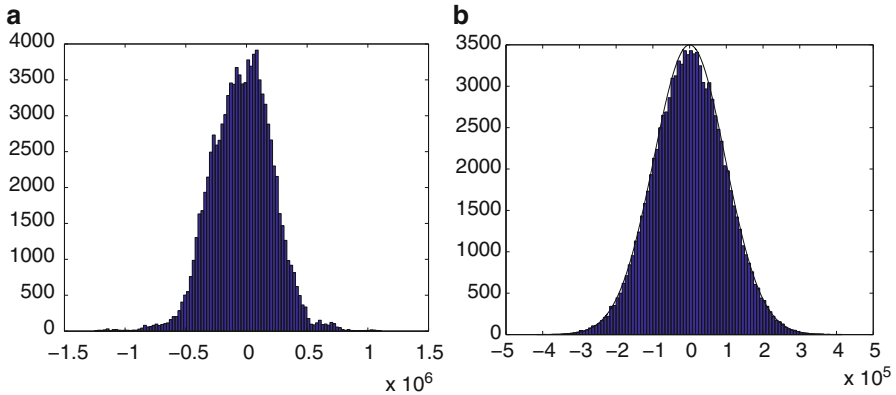
$$x_k = - \sum_{i=1}^p b_i x_{k-i} + n_k \quad (54.1)$$

where  $n_k$  is Gaussian noise, and  $b_i$  ( $i = 1, 2, \dots, p$ ) are the autoregression coefficients where  $p$  is autoregressive order. So

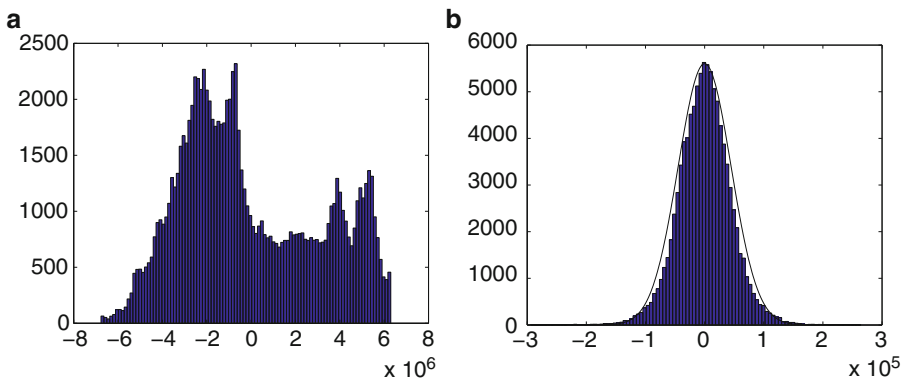
$$x_k - \sum_{i=1}^p c_i x_{k-i} = n_k \quad (54.2)$$

where  $c_i = -b_i$ . Which means the difference between samples (or their linear combinations) follows Gaussian distribution.

Based on the voice data we have collected, we observed that the vowel /a:/ samples ( $x_k$ ) don't follow Gaussian distribution, as illustrated in Fig. 54.1a, but when we choose  $p = 5$ ,  $c_1 = c_2 = c_3 = c_4 = 0$ ,  $c_5 = 1$ , i.e.,



**Fig. 54.1** (a) The histogram of 100,000 voice /a:/ samples  $x_k$ ; (b) the histogram of transformed voice samples  $x_k - x_{k-5}$  and its matching to a Gaussian distribution



**Fig. 54.2** (a) The histogram of 100,000 voice /a:/ samples  $x_k$ ; (b) the histogram of transformed voice samples  $x_k - x_{k-1}$  and its matching to a Gaussian distribution

$$x_k - x_{k-5} = n_k^a \tag{54.3}$$

the new sequences follow Gaussian distribution, as illustrated in Fig. 54.1b. Similarly, we observed that the vowel /i:/ samples ( $x_k$ ) don't follow Gaussian distribution, as illustrated in Fig. 54.2a, but if we choose  $c_1 = 1$   $p = 1$ , i.e.,

$$x_k - x_{k-1} = n_k^i \tag{54.4}$$

follows Gaussian distribution, as illustrated in Fig. 54.2.

We, therefore, tried to model the the new transformed voice sequences  $n_k^a$  and  $n_k^i$ , to see if a Gaussian MF can match its nature. For  $n_k^a$  and  $n_k^i$  from each subject (human) for 100,000 samples, we equally separate it into ten segments, and computed the mean  $m_i$  and std  $\sigma_i$  of the  $i$ th segment,  $i = 1, 2, \dots, 10$ . We also computed the mean  $m$  and std  $\sigma$  of the entire sequence (100,000 samples). To see which value  $-m_i$  or  $\sigma_i$  varies more, we normalized the mean and std of each segment using  $m_i \square m$ , and  $\sigma_i \square \sigma$ , and we then computed the std of their normalized values,  $\sigma_m$  and  $\sigma_{std}$ . We observed that  $\sigma_m \gg \sigma_{std}$ . We conclude, therefore, that if the transformed voice samples of each segment (short range) of the voice samples are Gaussian distributed, then the transformed voice samples in an entire video trAff0054ic (long range) is more appropriately modeled as a Gaussian with uncertain mean. This justifies the use of the Gaussian MFs with uncertain means to model the transformed voice samples.

### 54.3 Identifying Polyps Patient Voice Using Short-Time Fourier Transform and Singular-Value Decomposition

STFT uses a slide window to determine the sinusoidal frequency and phase content of a signal as it changes over time. The STFT of the voices is a matrix, how to extract its information for throat polyps detection? We use singular-value decomposition (SVD). The SVD is an important factorization of a rectangular real or complex matrix, with many applications in signal processing and statistics. Applications which employ the SVD include computing the pseudoinverse, least squares fitting of data, matrix approximation, and determining the rank, range and null space of a matrix. Given  $P \square C^{N \times M}$  (assuming  $N > M$ ), and  $rank(P) = r \leq M$ . Determine a numerical estimate  $r'$  of the rank of the data sets matrix  $P$  by calculating the singular value decomposition

$$P = U[\Sigma 000]V^T, \quad (54.5)$$

where,  $U$  is an  $N \times N$  matrix of orthonormalized eigenvectors of  $PP^T$ ,  $V$  is an  $M \times M$  matrix of orthonormalized eigenvectors of  $P^TP$ , and  $\Sigma$  is the diagonal matrix  $\Sigma = diag(\sigma_1, \sigma_2, \dots, \sigma_r)$ , where  $\sigma_i$  denotes the  $i^{th}$  singular value of  $P$ , and  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$ . Using SVD, the STFT of voices could be diagonalized, and the diagonal values in  $\Sigma$  could be used to represent the speaker voice power decay in the frequency domain. Generally the  $\sigma_1$  is much higher than  $\sigma_2$ , and the decay from  $\sigma_1$  to  $\sigma_2$  somehow represent how a person could handle his voice freely. For illustration purpose, we plot the singular values ( $\sigma_i$ ) ( $i = 1, 2, \dots, 10$ ) in Fig. 54.4 for the two patients whose spectrogram were plotted in Fig. 54.3. Observe Fig. 54.4, the voice power decay rate, i.e.,  $Pd = \sigma_1 - \sigma_2$ , is higher for a normal person than that of a patient with throat polyps, which means that a normal person could handle his/her voices more freely (with higher power changes from one

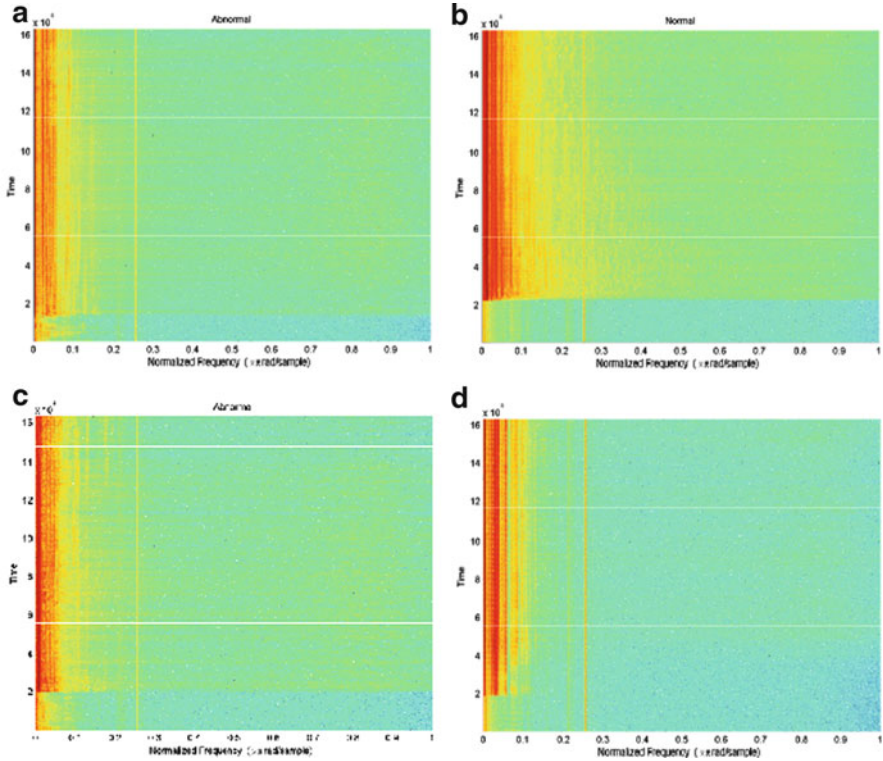


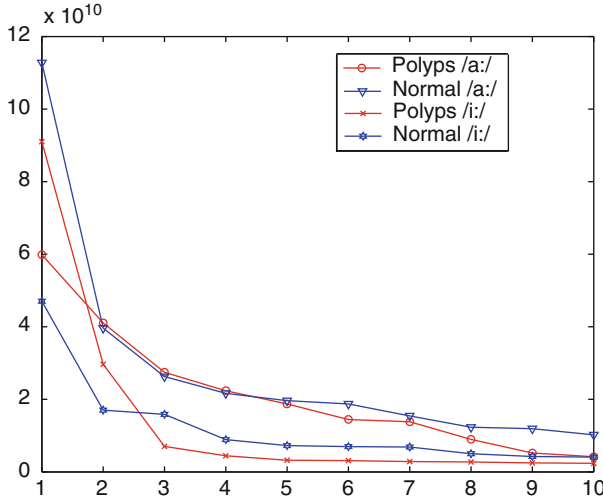
Fig. 54.3 The ten largest singular values of STFT of two patients in Fig. 54.4

frequency to another frequency). So voice power decay rate could be used as an identifier on throat polyps detection. In this paper, we will use the vowel /a:/ and /i:/ power decay rate in fuzzy classifiers for throat polyps detection.

### 54.4 Bayesian Classifier for Throat Polyps Detection

Bayesian decision theory [2] provides the optimal solution to the general decision-making problem. We assume that each patient has equal probability to have throat polyps, i.e.,  $H_1$ : Polyps, and  $H_2$ : Normal, so  $p(H_1) = p(H_2) = 0.5$ . If each transformed vowel voice samples (/a:/ and /i:/) of patient  $j$  follows Gaussian distribution,  $X_j \triangleq [x_j^a \ x_j^i]^T$  stands for the samples from patient  $j$  for vowel /a:/ and /i:/, then

$$p(\mathbf{x}_j|v_j) = \frac{1}{(2\pi)^{|\Sigma_j|/2}} \exp[-\frac{1}{2}2(\mathbf{x}_j - \mathbf{m}_j)^T \Sigma_j^{-1}(\mathbf{x}_j - \mathbf{m}_j)] \quad (54.6)$$



**Fig. 54.4** Spectrogram using a Short-Time Fourier Transform (STFT). Window size of STFT is 2,048, and overlap between two neighbor window is 1,024. (a) Is from a throat polyps patient speaking vowel /a:/; (b) is from a normal person speaking vowel /a:/; (c) is from the throat polyps patient speaking vowel /i:/; (d) is from the normal person speaking vowel /i:/

where  $m_j \triangleq [m_j^a \ m_j^i]^T$  and  $\Sigma_j = \text{diag}\{\sigma_j^{a2}, \sigma_j^{i2}\}$  are the mean vector ( $2 \times 1$ ) and covariance matrix ( $2 \times 2$ ) of  $\mathbf{x}_j$ . In this case,

$$p(\mathbf{x}|H_1) = \sum_{i=1}^{10} p(\mathbf{x}|v_i)p(v_i) \tag{54.7}$$

$$p(\mathbf{x}|H_2) = \sum_{i=11}^{20} p(\mathbf{x}|v_i)p(v_i) \tag{54.8}$$

Based on Bayes decision theory, since  $p(H_1) = p(H_2) = 0.5$ , we obtain the decision rule:

$$\text{Claim throat polyps if } p(\mathbf{x}|H_1) > p(\mathbf{x}|H_2) \tag{54.9}$$

$$\text{No throat polyps if } p(\mathbf{x}|H_1) < p(\mathbf{x}|H_2) \tag{54.10}$$

$$\text{Notsure if } p(\mathbf{x}|H_1) = p(\mathbf{x}|H_2) \tag{54.11}$$

This Bayesian polyps detector will be used in Sect. 54.5.

## 54.5 Simulations

We extract the general features and behavior of /a:/ and /i:/ voices for 20 patients, of which 10 have throat polyps and 10 have no throat polyps, and determine one discriminant rule for each patient in the domain of interest. In choosing the antecedents of the fuzzy classifier, we make full use of the statistical knowledge (mean and std) obtained from the patient voices. We used 100,000 samples in vowel /a:/ and /i:/ respectively to establish a discriminant rule for each patient. All-in-all, we obtained 20 rules, one per patient.

To evaluate the performance of the two fuzzy detectors, we used another group of 20 patients (testing group), which has no overlap with the first group of 20 patients whose vowel samples were used for fuzzy rules. By this means, it would help to demonstrate that our classifiers are robust. We also collected 100,000 voice samples for /a:/ and /i:/ respectively for each patient in the testing group. To demonstrate that our classifiers are able to detect throat polyps using a small number of samples, we made our detection based on every 5,000 samples, with 20 independent detections ( $20 \times 5,000$ ) for each patient. During testing, we obtain the mean  $\mathbf{m}^t = [m_a^t, m_i^t]$  for each 5,000 /a:/ and /i:/ samples.

### 54.5.1 Design of three Throat Polyps Detectors

#### 54.5.1.1 Design of Type-1 Fuzzy Polyps Detector

For a type-1 fuzzy classifier, the  $l$ th rule,  $R^l$ , is ( $l = 1, \dots, 10$ ):

$R^l$ : IF the transformed /a:/ voice is  $F_1^l$  and the transformed /i:/ voice is  $F_2^l$  and /a:/ power decay rate is  $F_3^l$  and /i:/ power decay rate is  $F_4^l$  THEN this patient has throat polypus (+ 1) [or throat normal (- 1)].

The antecedents  $F_k^l$  ( $k = 1, 2, 3, 4$ ) are described by a type-1 Gaussian MF whose mean,  $m_p^l$ , and std,  $\sigma_p^l$ , are determined by known patient voice samples. More specifically,  $m_1^l$  and  $\sigma_1^l$  are the mean and std of voice /a:/ samples in the 100,000 samples of patient  $l$  in the first group;  $m_2^l$  and  $\sigma_2^l$  are the mean and std of /i:/ samples in the 100,000 samples of patient  $l$  in the first group. To determine  $m_3^l$ ,  $\sigma_3^l$ ,  $m_4^l$ , and  $\sigma_4^l$ , we partition the voice samples into ten segments for /a:/ and /i:/ respectively, and obtain the STFT of each segment. Then apply SVD to the STFT matrix to obtain the power decay rate for each segments. The mean and std of the ten power decay rates are  $m_3^l$  ( $m_4^l$ ) and  $\sigma_3^l$  ( $\sigma_4^l$ ). The consequent corresponds to  $y^l = +1$  (polypus) or  $y^l = -1$  (normal) in the fuzzy detector.

For a type-1 fuzzy detector, its input,  $\mathbf{m}^t = [m_a^t, m_i^t, Pd_a, Pd_i]$ , is obtained from 5,000 vowel samples from patient in the testing group.  $Pd_a$  and  $Pd_i$  are the power decay rate for /a:/ and /i:/.

### 54.5.1.2 Design of Type-2 Fuzzy Polyps Detector

For type-2 fuzzy classifiers, the  $l$ th rule,  $R^l$ , is ( $l = 1, \dots, 10$ ):

$R^l$ : IF the transformed /a:/ voice is  $\tilde{F}_1^l$  and the transformed /i:/ voice is  $\tilde{F}_2^l$  and /a:/ power decay rate is  $F_3^l$  and /i:/ power decay rate is  $F_4^l$  THEN this patient has throat polypus (+ 1) [or throat normal (- 1)].

The antecedents  $\tilde{F}_k^l$  ( $k = 1, 2, 3$ ) are described by a type-2 MF, i.e., a Gaussian MF with uncertain mean, whose mean  $m_k^l \square [m_{k1}^l, m_{k2}^l]$  and std  $\sigma_k^l$  are determined by the voice samples of patients in the first group.  $F_3^l$  and  $F_4^l$  are same as those in type-1 fuzzy detector.

More specifically,  $\sigma_k^l$  ( $k = 1, 2$ ) are determined using the same method as described in Sect. 54.5.1.1, and  $m_{k1}^l$  and  $m_{k2}^l$  are determined as follows. We divided the 100,000 frames of the  $l$ th known patient into 10 equal-length (10,000 samples) segments, and computed the mean  $m_1^{lj}$  of /a:/ samples in the  $j$ th segment ( $j = 1, \dots, 10$ ). Let

$$m_{11}^l = \min_{j=1, \dots, 10} m_1^{lj}$$

$$m_{12}^l = \max_{j=1, \dots, 10} m_1^{lj}$$

so  $[m_{11}^l, m_{12}^l]$  is the range of uncertain mean of /a:/ voice samples of the  $l$ th known patient. We obtained the ranges of uncertain mean of /i:/ samples ( $[m_{21}^l, m_{22}^l]$ ) in a similar manner.

For a type-2 fuzzy detector, its input,  $\mathbf{m}^t = [m_a^t, m_i^t, Pd_a, Pd_i]$ , is obtained from 5,000 vowel samples from patient in the testing group.

### 54.5.1.3 Design of Bayesian Classifier

Observe from (54.6), that the Bayesian classifier needs  $\mathbf{m}_i = [m_i^a, m_i^i]^T$  and  $\Sigma_i = \text{diag}\{\sigma_i^{a2}, \sigma_i^{i2}\}$ . In our design,  $m_i^a$  and  $\sigma_i^a$  are the mean and std of vowel /a:/ in the 100,000 samples of patient  $i$  in the first group; similarly,  $m_i^i$  and  $\sigma_i^i$  are the mean and std of vowel /i:/ in the 100,000 samples of patient  $i$  in the first group; and, its input  $\mathbf{x} \triangleq \mathbf{m}^t$ , where  $\mathbf{m}^t$  is obtained from the mean value of 5,000 voice samples from a patient in the testing group.

## 54.5.2 Performance Analysis

We computed the average probability of miss detection ( $p_r(\epsilon)$ ) for each fuzzy detector as well as for the Bayesian detector in  $20 \times 20 = 400$  independent classifications (20 patients each with 20 5,000-sample segments), and please be



aware that the voices of the first group patients were used to design the fuzzy rules, and the testing group of patients have no overlap with the first group. Simulations show that  $p_r(\varepsilon) = 25\%$  for Bayesian classifier,  $p_r(\varepsilon) = 18\%$  for type-1 fuzzy classifier, and  $p_r(\varepsilon) = 14\%$  for type-2 fuzzy classifier.

## 54.6 Conclusions

Based on human voice samples and Hidden Markov Model, we showed that transformed voice samples (linearly combined samples) follow Gaussian distribution, further we demonstrated that a type-2 fuzzy MF, i.e., a Gaussian MF with uncertain mean, is most appropriate to model the transformed voices samples. We also applied STFT and SVD to the vowel voice samples, and observe that the voice power decay rate could be used as an identifier in throat polyps detection. Two fuzzy classifiers and a Bayesian classifier were designed for throat polyps detection based on human vowel voices /a:/ and /i:/ only, and the fuzzy classifiers are compared against the Bayesian classifier. Simulation results showed that an interval type-2 fuzzy classifier performs the best of the three classifiers.

## References

1. de Oliveira Rosa M, Pereira JC, Grellet M (2000) Adaptive estimation of residue signal for voice pathology diagnosis. *IEEE Trans Biomed Eng* 47(1):96–104
2. Duda RO, Hart PE (1973) *Pattern classification and scene analysis*. Wiley, New York
3. Rabiner LR (1989) A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE* 77(2):257286