Research and Implementation of an UAV Engine Debugging Earmuff System

Changliang Liu 1,*, Dongsheng Lin 1, Xing Zheng 1, Shaojun Liang 1, Zhongzhu He 1, Haitao Zheng 1, Lanqiong Sun 1 and Xing Su 2

1 Army Engineering University Non-Commissioned Officer College, Wuhan, China
2 Unit 32151 of PLA, Xingtai, China

*Corresponding author e-mail: lclear@126.com

Abstract. Aiming at the problem that speech communication can not be implemented due to the huge noise of UAV engine debugging, a debugging earmuff system based on radio communication is designed. The system captures the debugger's mouth shape by camera, and uses lip recognition technology to obtain the instructions expressed by the debugger through region detection, lip feature extraction, training and recognition. After modulation, the instructions are transmitted to the distant place by the wireless transmitter module in the form of electromagnetic waves. At the executing end, the antenna is used to receive the radio information and to release it by voice after demodulation, so that the cooperator receives the instructions from the debugging end.

1. Introduction

At present, the development of UAV is changing rapidly, many of them use piston two-stroke or four-stroke aero-engines. In order to ensure the UAV to carry out its mission smoothly, the UAV must check and adjust the working state of the engine before take-off. The noise of the engine starting and running is huge, and the UAV can not communicate with the language within tens of meters. However, the debugger needs to change the running state of the engine in time to obtain various working conditions of the engine. The following parameters require corresponding instructions to the main control station. Because there is no dialogue at the engine, the traditional way is to gesture to the cooperating personnel in the distance, understand the intention of the debugging personnel through the agreed gesture in advance, and then give the debugging personnel the desired instructions to the main station. For example, it is agreed that if the engine is to be turned to high horsepower, the Commissioner will repeat the movement upward with the palm of his hand. When the cooperator observes the movement, he knows that the command of "high horsepower" needs to be issued. Then, by issuing the command of "high horsepower" to the main station on the intercom, the engine will be turned to high horsepower. This process has many links, low efficiency and easy to distract the debugger's attention. The consequences of distraction are often very serious. It can involve propellers in clothing, or even kill people at high speed.

Aiming at the serious problems existing in engine debugging, this paper provides an unmanned aerial vehicle engine debugging earmuff system. To solve the problem that traditional UAV engine debugging can only convey its intention by gesture, which not only has many links, but also is inefficient and easily distracts the debugger's attention, thus causing serious casualties.
2. General design idea
A Through the analysis, it is found that the reason for the difficulty of debugging the traditional UAV engine is that the information conveyed by the voice is submerged in the high noise environment. As a result, the sound signal can not be transmitted. It can only be transmitted by visible light, that is, by using gesture distinction and human eye observation. In view of the fatal shortcomings mentioned above, we may consider replacing the information transmission media. Nowadays, wireless communication develops rapidly and technology is mature. It can transmit information completely through electromagnetic wave. What we need to do is to optimize the transmitter and receiver so as to make them easy to use, compact and efficient.

![Figure 1. General structure of the system](image)

3. Software and Hardware Construction of the System

3.1. Hardware Structure of Debugging End
The most reasonable design of the debugger's end is to make headwear earmuffs. On the one hand, they are easy to wear. On the other hand, they can provide protection for the debugger's hearing.

On this basis, the acquisition of debugging instructions has become a crucial part. As mentioned above, no sound signal can be used in high noise environment. Visible light is the only way. When the debugger's mouth sounds, besides the change of sound, there is also a change of mouth shape. Therefore, when the voice is not available, we can use the information conveyed by the change of mouth shape.

Based on the above analysis, a camera is equipped on the debugging earmuff to capture the image of the debugger's lip, and then the command signal expressed by the debugger can be obtained by identifying and analyzing it. The schematic diagram of the overall structure of the debugging earmuffs is shown in Figure 2.

![Figure 2. Overall structure diagram of unmanned aerial vehicle engine debugging earmuffs](image)

1-right earmuff, 2-first receiving antenna, 3-connection band, 4-first transmitting antenna, 5-left earmuff, 6-universal tube, 7-camera

The UAV engine debugging earmuffs include right earmuffs, receiving antennas, connection bands, transmitting antennas, left earmuffs, universal tubes and cameras.
The concrete structure is that the right earmuff is connected with the left earmuff through the connecting belt, one end of the universal tube is connected with the left earmuff, the camera is set at the other end of the universal tube, the receiving antenna is set on the right earmuff, and the transmitting antenna is set on the left earmuff.

As shown in the figure above, the right loudspeaker, the signal recognition processing board and the battery are all arranged in the right cover body, and the right sponge coil is arranged on the right cover body. The battery connects the signal recognition processing board with power supply, the receiving antenna is connected with the signal recognition processing board, the data input terminal of the signal recognition processing board is connected with the camera, and the data output terminal of the signal recognition processing board is connected with the right loudspeaker and the left loudspeaker in the left earmuff; the signal recognition processing board is connected with the microcontroller STM32F3 and the image. The acquisition chip SAA7111 and the wireless transceiver chip nRF24L01 are composed.

The working block diagram of the whole debugger side is shown in Figure 3.

**Figure 3.** Functional block diagram of unmanned aerial vehicle engine debugging earmuff circuit

The specific working process is as follows: Firstly, the image acquisition chip SAA7111 on the signal recognition processing board captures the monitoring image from the camera and sends it to the microcontroller STM32F3 for recognition processing. The processing result is sent to the wireless transceiver chip nRF24L01 by the transmission antenna. The external wireless instruction signal is received by the receiving antenna and sent to the wireless transceiver chip nRF24L01 for decoding, and then sent to the microcontroller STM32F3 for processing to form an audio signal, which is output to the loudspeaker and heard by the debugger's ears.

### 3.2. Functional Implementation of Debugging End

After the above analysis, the key technology of debugging end is lip image recognition and processing based on STM32F3. It is divided into the following steps to finish.

#### 3.2.1. Labial Region Detection

This is the first step. Its accuracy is directly related to whether lip movement features can be extracted effectively, thus affecting image recognition and processing. At present, most research institutes follow the principle of lip localization: first, face detection, and then lip localization in the face area. However, this paper is headphone style and does not need to photograph the whole face. Therefore, this paper adopts a color-based method.

The chromaticity distribution and chromaticity characteristics of lips were analyzed. The effective lip regions were detected by color space conversion and pseudo-color enhancement.

Red Exclusion proposed by Lewis [20] et al. is a classical method based on color detection of lip region. This method does not need to change the color space. It considers that the main color of face and lip is red. In view of the different proportion of red in lip and skin color, the red component is removed, and the difference of lip color and skin color is expressed by blue and green components. The discriminant is shown in the following formula.

\[
\log(G/B) < \beta
\]
Among them, G and B represent the green and blue component values respectively, and beta is the threshold value calculated by statistics. The pixels satisfying the above formula are lip color pixels, otherwise they are skin color pixels.

3.2.2. Lip feature extraction. Feature extraction is a key link in lip reading system. Whether the extracted feature vectors are representative or not will directly affect the recognition rate of lip reading. Current lip feature extraction methods can be divided into three categories: shape-based method, pixel-based method and blending method. Pixel-based method considers that all the pixels in ROI (Region Of Interest) region represent the visible voice information together. For ROI region, a series of feature vectors are obtained by transforming and reducing dimensions (such as PCA, DCT, LDA, DWT, etc.) to represent lip movement characteristics. In order to ensure the fast and effective image processing, we often need to transform the image into the frequency domain for processing. Discrete Cosine Transform (DCT) converts the original image information block into a set of coefficients representing different frequency components with its fast algorithm, which makes the description of unimportant components require only a small number of bits. In addition, the frequency domain decomposition also maps with the processing of human visual system.

For the set of coefficients obtained by DCT transform, the Zigzag scanning method [1] takes the upper left corner data first, retains the low-frequency components with high energy, discards most of the high-frequency coefficients, and then reduces the dimension. PCA algorithm is used to select the eigenvectors corresponding to the first n larger eigenvalues by using the contribution rate of eigenvalues to form the feature space. K. means algorithm in clustering algorithm is used to divide the n-dimensional vectors of these rows into several different categories, and each n-dimensional vector is allocated to a clustering center as an input variable for subsequent recognition.

3.2.3. Training and Recognition. Hidden Markov Model (HMM) is established, which is denoted by five coefficients as $\lambda = (M, N, \pi, A, B)$, usually abbreviated as $\lambda = (\pi, A, B)$. The HMM model is built for every word (or word) in the database by sample sequence. In the training stage, the parameters $M, N, \pi, A, B$ are initialized first, and then the final parameters of the model are estimated iteratively using forward-backward algorithm. Viterbi algorithm is used in the recognition process, and the process is as follows:

1. The feature vectors are obtained by lip detection and feature extraction.
2. For each HMM model $\lambda, \lambda, \ldots, \lambda$, Viterbi algorithm is used to calculate the probability $P(V | \lambda)$ of generating feature vectors to be identified, and the maximum probability is obtained by comparison.
3. The word (or word) corresponding to the model $\lambda$ with the maximum probability $P$ is the recognition result.

3.2.4. Instruction Sending. Recognizing the debugger's mouth type means that the instructions issued by the debugger have been formed and stored in Chinese characters. At this time, the Chinese characters of the instructions can be converted into audio signals by voice conversion module, which can be loaded into the wireless communication module and transmitted through the transmitting antenna, waiting for the receiving device to receive.

3.3. Hardware Architecture of Executor

The executor is a device which is placed in the main control station to transmit debugging instructions to the controller. It receives radio information from the debugging end and releases it by voice. After that, it should be able to transmit the command return of the control station personnel to the debugging personnel. The specific way is to pick up the voice signal by microphone and transmit it by the corresponding module. Its concrete structure is shown in the Figure 4. The circuit function block diagram is shown in Figure 5.
Figure 4. Overall structure diagram of UAV engine debugging earmuff operator end
16-second transmitting antenna, 17-second receiving antenna, 18-microphone, 19-loudspeaker, 20-
matching device host

Figure 5. Circuit function block diagram of matching device

The UAV engine tuning earmuff device also includes a second transmitting antenna 16, a second
receiving antenna 17, a microphone 18, a large loudspeaker 19 and a host 20. A second signal
recognition processing board is set in the host 20, and a wireless transceiver chip nRF24L01 connected
with the second microcontroller STM32F3 and the second microcontroller STM32F3 is set on the
second signal recognition processing board. The second transmitting antenna 16 and the second
receiving antenna 17 are respectively connected to the wireless transceiver chip nRF24L01, the
microphone 18 and the large one. The speaker 19 is connected to the second microcontroller STM32F3.
The second receiving antenna 17 receives the wireless signal and sends it to the second wireless
transceiver chip nRF24L01 for decoding, and then transmits it to the second microcontroller STM32F3
for processing to form an audio signal output to the loudspeaker 19. The microphone 18 is responsible
for collecting feedback voice instructions and sending them to the second microcontroller STM32F3
for processing to form a digital signal through the second wireless transceiver. The chip nRF24L01 is coded
and transmitted by the second transmitting antenna 16.

3.4. Functional Implementation of Executor
Executor is a typical radio circuit. The radio wave with command information can be received by the
second receiving antenna of the executor after it is transmitted from the debugger. The command is
converted into audio signal from the loudspeaker through a series of links such as tuning, frequency
conversion, IF amplification, demodulation, low frequency amplification and power amplifier.
4. Overall operation mode
When the UAV engine is debugging earmuffs, under the control of the universal tube, the camera accurately monitors the mouth shape of the debugger, detects the lip area by the signal recognition processing board, and then carries out image transformation to obtain data. After dimensionality reduction, the final feature vector is obtained by vector quantization to extract features, and through the event. The input type-voice relationship matches and identifies the voice instructions to be expressed by the exit type, and then is transmitted by the wireless transceiver chip nRF24L01 through the first transmitting antenna.

As shown in Fig. 4 and Fig. 6, the second receiving antenna receives the above voice instructions and sends them to the host signal recognition processing board to restore the electrical signals to form voice instructions to play through the loudspeaker; the cooperator executes the instructions immediately after hearing them, and then sends out the return instructions, which are picked up by the microphone and sent to the host signal. The identification processing board is processed and then sent by the second wireless transceiver chip nRF24L01 through the second transmitting antenna.

Then, as shown in figs. 1, 2, 3 and 5, the first receiving antenna receives the aforementioned instruction return and is processed by the first signal recognition processing board and sent to the right speaker and the left speaker to form the instruction return speech, which is transmitted to the debugger's ear.

5. Conclusion
Compared with the existing means, the system has the following advantages:
1) Solve the problem that voice can not be used to transmit information in high noise environment;
2) Streamline the transmission link of debugging information to the main station, which is more efficient;
3) Realize direct communication between the controller and the debugger, which is more direct;
4) Provide hearing protection while guaranteeing communication.

References
[1] Qiuqi Ruan. Digital Image Processing [M]. Beijing: Electronic Industry Publishing House, 2013:75-76.
[2] M.Heckmarm,etc.DCT-based video features for audio-visual speech recognition. Proc. International Conference of Spoken Language Process. Denver, USA: September, 2002: 1925-1928.
[3] L.R.Rabiner.A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. Proceedings of the IEEE,1989,77(2).
[4] He Jun.Research on Speech and Lipreading Human Machine Interaction for Service Robot[D]. Nanchang: Nanchang university, 2009.