

Military Technical College  
Kobry Elkobbah  
CAIRO - EGYPT



8<sup>th</sup> International Conference  
on Aerospace Sciences &  
Aviation Technology

## Proposed System for Securing Aircraft Guidance Commands

*Assco. Prof. Dr. Elsayed Soleit*

*Assco. Prof. Dr. Mohmoud Gadallah*

*Eng. Ashraf Mahran*

*Egyptian Armed Forces*

### Abstract

This paper introduces a system for transmitting the aircraft guidance command over a narrow band communication channel in a secured way. The paper takes into account the noise, which exists in the guidance center. The proposed system is based on speech recognition and consists of two modules; one on the ground and the other is on the onboard. Guidance command words are modeled using Hidden Markov Modeling (HMM) and stored in the onboard module. During operation the input guidance word is processed and transmitted in a coded form from the ground module to the aircraft. The onboard module recognizes the words using HMM speech recognizer. The feature extraction technique used in this work is the Perceptual Linear Predictive (PLP) coding, which shows good immunity to noise. The performance measure is the recognition accuracy at different SNRs, as well as the achieved compression ratio of the transmitted message.

### 1 Introduction

The rapid development in speech recognition technology opens new areas of applications. Recently, these applications have taken place in military environment. The speech processing technique is used to transmit the speech signal, in coded form, through a narrow band channel; reducing the required bit rate. Hence, using this criterion for securing the aircraft guidance commands is of great benefits.

A proposal for a system to secure aircraft guidance commands using speech recognition is depicted in Fig.1 that can be used in the guidance center for ground-to-air communication. A similar system can be used in the aircraft for air-to-ground communication. The proposed system includes two principle modules; the ground module and the onboard one. The ground module is equipped in the guidance center where the input speech signal that represent the guidance vocabularies (heading, bearing, altitude...) is applied to a preprocessing unit. The

function of the preprocessing unit is to determine the end point detection, blocking the speech signal into frames, and speech signal pre-emphasis. Features of the speech signal are extracted in feature extraction unit. A set of information indices representing the spoken words is generated from the vector quantization process. These indices are transmitted to the aircraft through a data transmission channel. On the onboard module the received indices are the input to HMM recognizer to obtain the output command word.

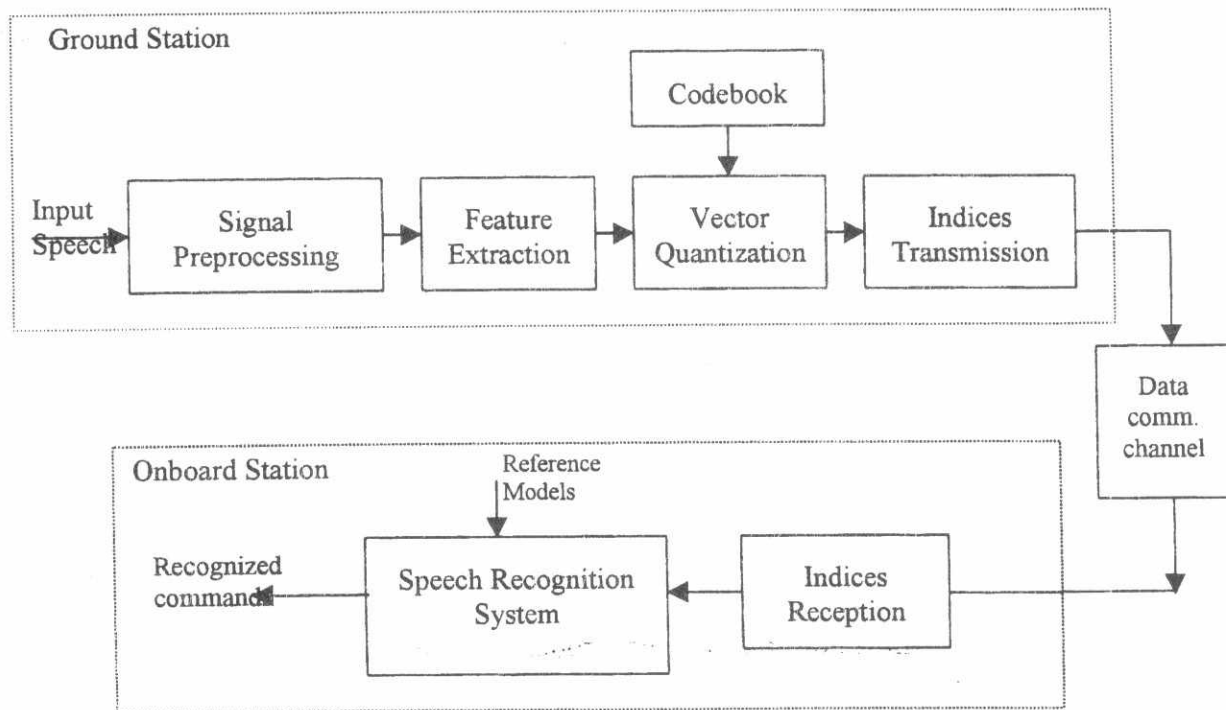


Fig.1 A Proposed System for Securing Aircraft Guidance Commands

## 2. Signal Preprocessing

The speech signal via a microphone is sampled and converted into digital formats using the sound blaster that includes 8 bits A/D.

Rabiner's algorithm for End-Point Detection (EPD) [1] is used to remove the background noise before and after speech interval. This technique is based on energy and zero-crossing rate to discriminate between noise and speech signal.

The speech signal is divided into frames. Each frame is of length 13.5 ms. and successive frames are overlapped by 0.5 frame length. Each frame is weighted using Hamming window.

The speech signal is passed through a first order high pass filter (pre-emphasis filter) to limit noise and to enhance the higher frequency components.

### 3. Feature Extraction of Speech

#### *Perceptual Linear Prediction Coding of Speech*

The PLP technique incorporates three basic concepts about the human hearing system [2]. These three basic concepts are the critical band resolution curve, the equal-loudness curve, and the intensity-loudness power-law relationship. These three physical components of the hearing process, combined with an autocorrelation model of the speech signal help to create a better speech coding technique with immunity to noise.

PLP can be summarized as follows [3]:

1. The power spectrum of each frame is calculated and warped onto a Bark scale using Eq.1:

$$\Omega(\omega) = 6 \ln \left\{ \left( \frac{\omega}{1200\pi} \right) + \sqrt{\left( \frac{\omega}{1200\pi} \right)^2 + 1} \right\} \quad (1)$$

2. The speech signal spectrum is filtered using successive critical band filters in Eq.2, which simulates the frequency resolution of the ear:

$$CB(\Omega) = \begin{cases} 10^{(\Omega+0.5)} & \text{for } \Omega \leq -0.5 \\ 1 & \text{for } -0.5 < \Omega < 0.5 \\ 10^{-2.5(\Omega-0.5)} & \text{for } \Omega \geq 0.5 \end{cases} \quad (2)$$

3. The Bark scale spectrum is compensated for the non-equal perception of loudness at different frequencies by the filter given in Eq.3:

$$E(\omega) = \frac{\omega^4 (\omega^2 + 56.8 \times 10^6)}{(\omega^2 + 6.3 \times 10^6)(\omega^2 + 0.38 \times 10^9)(\omega^6 + 9.58 \times 10^{26})} \quad (3)$$

4. The nonlinear relation between a sound's intensity,  $I$ , and its loudness,  $L$ , is applied by Eq.4:

$$I = \sqrt[3]{L} \quad (4)$$

5. The Inverse DFT of the filtered power spectrum is calculated. Then the PLP coefficients for the speech segment are calculated using Durbin's algorithm.

### 4. Codebook Design and Vector Quantization

#### *(i) Code-Book Design*

Using Linde-Buzo-Gray algorithm [4], a 512 codebook vectors are designed from 17426 feature vectors randomly distributed in space, representing the training data.

The distortion measure is chosen to be the nearest neighbor based on Euclidean distance given by:

$$d^2(x, y) = \sum_{i=1}^L (x_i - y_i)^2 \tag{5}$$

where  $x_i$  and  $y_i$  are the  $i^{th}$  components of the  $x$  and  $y$  respectively and  $L$  is the number of coefficients for each feature vector.

**(ii) Vector Quantization**

The reference 512 vectors from the codebook design algorithm are stored in memory. The feature vector under consideration is assigned an index according to the nearest neighbor rule mentioned in distortion measure. After this process, each one-second of speech could be only stored in only 0.225 Kbytes instead of 22.05 Kbytes. These indices will be used in training the HMM during training phase. While during operation (i.e. recognition phase) these indices will be transmitted to the aircraft as a low bit rate signal, which needs narrow bandwidth channel.

**5. Speech Recognition System**

**• Training the HMM Recognizer**

The training phase can be summarized in the following steps [5]:

1. Read the observations of the words, which are used for building the model. These observations are a set of indices resulting from vector quantization unit. Each word is repeated five times and end-point detection is done manually.
2. Initialize the model according to the following:

The state transition probability matrix  $A$  as:

$$A = \begin{bmatrix} 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 0 & 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\ 0 & 0 & 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 0 & 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{6}$$

Also, the observation symbol probability matrix  $B$  as:

$$B = \begin{bmatrix} 1/M & 1/M & \dots & 1/M \\ 1/M & 1/M & \dots & 1/M \\ 1/M & 1/M & \dots & 1/M \\ 1/M & 1/M & \dots & 1/M \\ 1/M & 1/M & \dots & 1/M \\ 1/M & 1/M & \dots & 1/M \end{bmatrix} \quad (7)$$

And the initial model distribution as:

$$\pi = [1 \ 0 \ 0 \ 0 \ 0 \ 0] \quad (8)$$

3. Calculate the probability  $P(O/\lambda)$  using the forward and backward algorithms where  $O$  is the observation sequence and  $\lambda_i$  represents the HMM model parameters  $(A, B, \pi)$ .
4. Using Baum-Welch algorithm to find new estimate values of  $A_{est}$ ,  $B_{est}$ ,  $\pi_{est}$ .
5. Calculate the likelihood of the model, which is given by:

$$P(O/\lambda_{est}) = \prod_{x=1}^L P_x \quad (9)$$

$$P_x = P(O^x / \lambda_{est})$$

where  $L$  is the number of the words in the training data of each word which is five words.

5. Calculate the distance,  $dis$ , which is given by

$$dis = \frac{P_i - P_{i-1}}{P_{i-1}} \quad (10)$$

where  $P_i$  is the present likelihood and  $P_{i-1}$  is the previous likelihood. This value is compared with a threshold value  $\epsilon$ . If it is larger, the procedure 3-6 are repeated until  $dis$  is less than  $\epsilon$  which is chosen to be 0.0005.

• **Transmitting the Indices**

According to Fig.1, the speech signal after being processed in the ground module; a set of indices representing the spoken message is obtained. These indices are the output from vector quantization process ranging between 1:512. These indices will be transmitted to the aircraft recognition system through a data transmission channel to be displayed on A/C display. In this work we assume an ideal channel.

### • The HMM Recognizer

According to Fig.1, the received indices after being recovered from the receiver unit are applied to the recognizer unit. The steps for the recognition process are as follow [5]:

1. Read the observation sequence,  $O = \{o_1 o_2 \dots o_T\}$ , of the transmitted word.
2. Start with the first reference model  $(A, B, \pi)$ ; compute the probability of this word being generated by this model,  $P(O/\lambda)$ , using Alternative Viterbi algorithm.
3. Repeat the step 2 for all the stored models. The unknown transmitted word is assumed to be generated from the model that gives the highest probability.

## 6. Improving Recognition Accuracy in Noise

The problems that associate the application of the proposed system in securing the aircraft guidance commands are [6]:

- (i) The noisy environment of the guidance center, in which the ground part is used.
- (ii) The noise of the channel.
- (iii) The noise of the aircraft board if the transmitted message is required to be heard.

Problem (iii) can be overcome, simply, in this system by allowing the system to display the text of the recognized command or hearing it using speech synthesis. In this work, we focus our attention to the problem of the noise in the guidance center. To study this problem, we consider a noise free channel as mentioned above.

It is known that the accuracy of speech recognition systems degrades rapidly in the presence of noise and distortion. According to Fig.1, the ground station is responsible for generating the guiding commands of the aircraft. However, the environment of the ground station is not noise free. Unfortunately, there are many sources of noise (such as PC's, printers, teletypewriter) and we can't anyway work without either of them.

We investigate the effect of noise on the recognition accuracy of our system, which consists of 40 word represents the commonly used words in guidance management. The system assumed to be trained with clean words (noise free conditions). Fig.4 demonstrates the dependence of the recognition accuracy on the variation of the input SNR for a clean model.

It is apparent that as the input test words are noise free, the system verifies high recognition accuracy. As the input test words are corrupted with an additive noise, the recognition accuracy varies. The recognition accuracy degrades significantly as the input SNR decreases below 30 dB.

From our previous work in [7] we found that by contaminating the training data and by using a noise immune feature (PLP) we can obtain better recognition accuracy in noisy environment. Moreover, It is found that the system can perform approximately the same recognition accuracy for a margin of SNR round the SNR of the reference model.

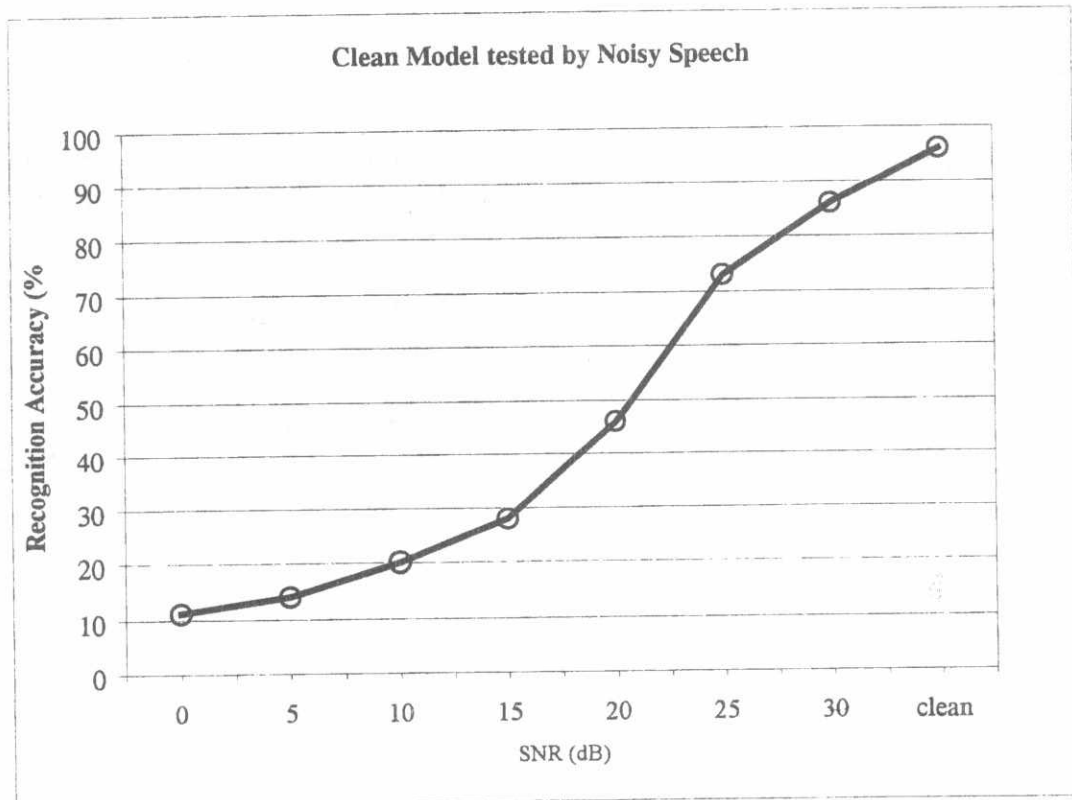


Fig.4 Recognition accuracy versus different SNR

## 7. Simulation Results

The operation of the proposed system can be summarized as follows:

The controller in ground center is going to send his message to the pilot. He is going to speak via a microphone. This signal is then processed and coded via a coding technique and then transmitted to the aircraft.

At the A/C this coded speech signal is compared with the stored reference models of speech to recognize the corresponding model. The recognized message is then displayed on a display so that the pilot can read it easily.

For a clean model tested with noise free speech, the system gives an accuracy of 96%. It was found that sources include office machinery, *as in guidance centers*, such as a typewriters or printers, personal computers, or workstations which are usually equipped with moving components like disks and fans, telephone ringing and background conversation of other people. The SNR is ranging from 15 to 20 dB. When we model our system with a 15 dB noisy signal it shows an accuracy of 92%.

As shown from Fig.5, which shows the recognition accuracy of a 15-dB model, tested with noisy speech with SNR from (9 dB to 21 dB). Maximum recognition accuracy is achieved when matching occurs and as SNR of test data differs from reference one; recognition

accuracy decreases. The system can approximately perform the same recognition accuracy in the range (13 dB to 19 dB).

The achieved compression ratio in data transmission rate is 1:100, which results from transmitting an index of each frame (9 bits) instead of 256 samples and each sample is 8 bits.

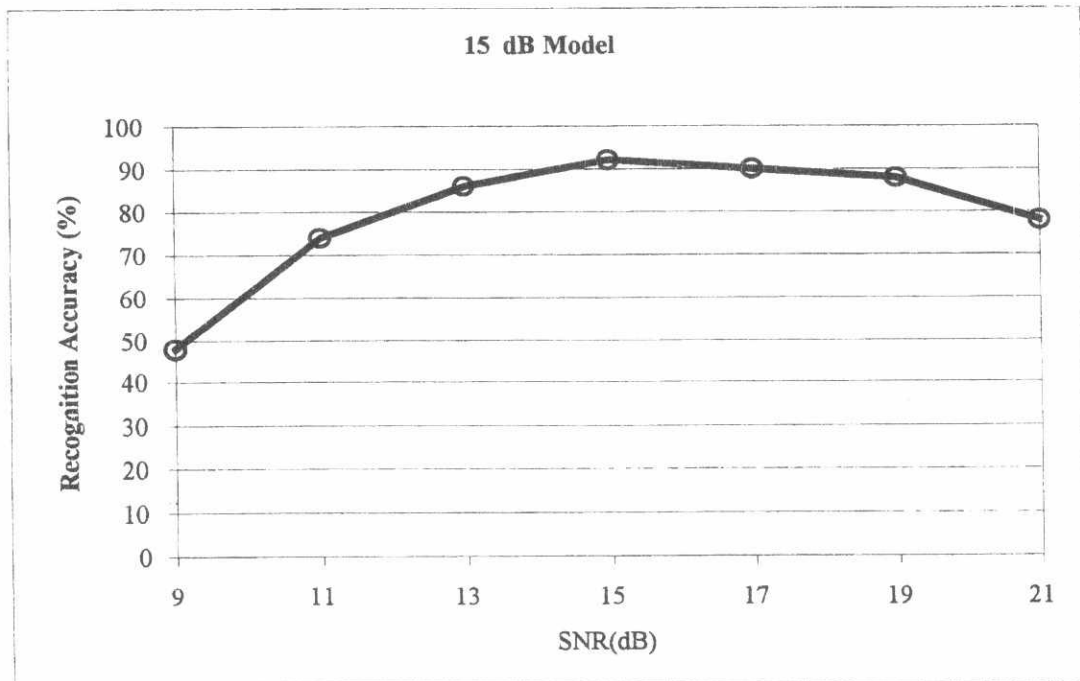


Fig.5 15-dB model tested with noisy data

### 7. Conclusions and Recommendations

The proposed system has the following advantages:

- Reduction in transmission rate via transmitting the codebook indices, for example a 10 minutes speech message can be send in 6 seconds only which minimizes the time at which the channel is on.
- Security by transmitting the coded signal (indices).
- It reduces the working load to the pilot, which can reach 5 g.
- Many messages can be stored and redisplayed in any desired order.
- The pilot will be able to understand the message correctly because it will be displayed for a considerably sufficient time rather than the spoken message, which is heard just once. Also, messages can be accessed using database capabilities.



- At noise level in A/C cockpit, Sound Pressure Level (SPL) equals 90 dB or more, the speech signal is barely intelligible. So, receiving the speech signal in a text format is an advantage in this case.

Although reduction of transmitted data rate would result in reduction in error rate but a complete study of the channel is required in the future work as well as the noise at the aircraft cockpit in which air-to-ground link is equipped.

## References

- 
- [1] L. R. Rabiner and M. R. Sambur, "an algorithm for determining the End Points of Isolated Utterances", Bell system Technical Journal. Vol. 54, No.2 pp 297-315, 1975.
  - [2] C. Dobrin, P. Haavisto, K. Laurila, J. Astola, "Speech Recognition Experiments in a Noisy Environment using Auditory System Modelling", Eurospeech 95 V1.P131, 1995.
  - [3] Carnegie Mellon University CMU: Robust Speech Recognition Group (Subset of ARPA speech recognition community). <http://www.cs.cmu.edu>
  - [4] Kai-Fu Lee, "Large-Vocabulary Speaker-Independent Continuous Speech Recognition; The SPHINX System", Ph.D. Thesis, Carnegie Mellon University, 1988.
  - [5] L. R. Rabiner and B. Juang, "Fundamentals of Speech Recognition" Printice-Hall Inc., Englewood Cliffs, NJ 07632, 1993.
  - [6] A. Mahran, "Application of speech Coding Techniques for Securing Aircraft Guidance Commands", M.Sc. Thesis, Military Technical College, 1998.
  - [7] M. Gadallah, E. Soleit, and A. Mahran, "Noise Immune Speech Recognition System", URSI, Ain Shams Univ., 1999.