NINTH EUROPEAN ROTORCRAFT FORUM

Paper no. 44

# THE APPLICATION OF DIRECT VOICE CONTROL TO BATTLEFIELD HELICOPTERS

S. D. ROY, A. CORT, G. A. RIXON

WESTLAND HELICOPTERS LIMITED

YEOVIL, ENGLAND

September 13-15, 1983

STRESA, ITALY

Associazione Industrie Aerospaziali

Associazione Italiana di Aeronautica ed Astronautica

# THE APPLICATION OF DIRECT VOICE CONTROL TO BATTLEFIELD HELICOPTERS

S. D. Roy, A. Cort, G. A. Rixon

## ABSTRACT

The high workload experienced by battlefield helicopter crews is due to the requirement to fly at high speed at extremely low altitude, and simultaneously operate cockpit equipment such as radio sets, navigation aids, maps etc. Voice control offers the potential of providing a means by which the crew can operate the cockpit systems without either removing their hands from the flying controls or taking their eyes off the outside world.

The paper discusses the state of the art of voice control technology, its potential helicopter applications, the technical steps which must be made to implement voice control and the technical problems of performance measurement and specification.

Three key steps need to be made to make voice control practical in helicopters;

- o Development of recognition systems which operate with a high recognition probability in the noise levels found in the helicopter cockpit.
- o Development of efficient man/machine dialogues (good human factors).
- o Development of integrated cockpit systems capable of control by voice.

## 1.0 INTRODUCTION

Control of helicopter systems by voice offers considerable advantages over more conventional methods of control particularly in situations where the crews' hands and eyes are occupied with other tasks.

Over the last two decades, considerable effort has been spent by many research laboratories on developing computer speech processing techniques, to the extent that there are now available a large number of experimental systems of varying degrees of performance, some suitable for industrial applications and a few beginning to show the capabilities required for aircraft and helicopter applications (Refs: 2, 3 & 4).

At Westland, work is continuing on two fronts; firstly to specify the performance of advanced recognisers for cockpit applications that are capable of accepting more natural speech, have increased vocabulary and are tolerant to wider variations in an individual's speech. Secondly to develop operating procedures compatible with the tactical and operational requirements found in Battlefield Helicopters. Both activities are part of a wider programme to develop reduced-workload cockpits compatible with the next generation of helicopters.

## 2.0 CREW WORKLOAD IN A BATTLEFIELD HELICOPTER

### 2.1 The Battlefield Helicopter

Helicopters have several roles in the battlefield including;

- o Reconnaissance (as an observation point and for deploying patrols).
- o Fire support (anti armour).
- o Combat support (urgent transport).
- o Command/control (mobile command post).

All of these roles place similar severe requirements on the piloting of the aircraft, particularly for Tactical Low Level Flight (TLLF), to improve survivability when near the forward line of own troops (FLOT). These roles require a very high percentage of time to be spent flying "eyes out" both for obstacle avoidance and observation, the latter being necessary for navigation, reconnaissance, and because of the threat of attack.

### 2.2 Crew Workload

The workload distribution during various stages of a battlefield helicopter mission is shown in Fig. 1. The visual and manual channels of the crew are very heavily loaded, being virtually continuously occupied by piloting the aircraft and monitoring battlefield conditions.

Pilots are able to control their present aircraft systems whilst flying at low altitude by becoming sufficiently familiar with the control layout to operate most of the controls by touch. Although attempts have been made to ensure that all controls can be discriminated by touch or position, the increase in number and complexity of avionic systems will render this impossible. This same increase is also leading to a shortage in panel space, particularly given the helicopter requirement of good external vision.

One solution to the control problem is a centralised multifunction Control and Display Unit (CDU). This consists of a small alpha-numeric display and a keyboard. However, these cannot be operated by touch alone and require even more concentration on manual and visual operation than the conventional controls.

Thus conventional dedicated controls are tending to be replaced by a control which has the advantage of requiring much less panel space but which requires more "Hands Off" and "Eyes In" operation. With the high workload of low altitude flying at night, or TLLF during daytime, the operation of complex cockpit systems would simply become impractical.

### 2.3 Advantages of Direct Voice Input

Direct Voice Input (DVI) offers a solution to the lack of space for controls and allows the pilot to control the aircraft systems while flying "eyes out" and "hands on".

The use of DVI transfers some of the pilot's workload from the visual and manual channels to the audio channel. Fig. 1. shows how this might effect various tasks during each phase of a mission. These changes are particularly important during Tactical-Low-Level-Flight and hover and when piloting and weapons operation impose a high visual and manual workload. Although DVI will have very little direct impact on piloting and weapons operation, the transfer of avionics to audio control will leave the hands and eyes devoted to the primary tasks. With conventional controls the secondary tasks (i.e. communications and navigation) are likely to be compromised, by using DVI we will get significant improvements in the controllability of systems and hence the effectiveness and safety of the mission and aircraft.

## 3.0 DVI TECHNOLOGY

### 3.1 Types of Automatic Speech Recogniser

Speech recognition systems may be categorised on two parameters.

- o Speaker dependent or independent; i.e. whether or not the system has to be retrained by each user.
- o Isolated word, connected word and continuous speech, i.e. the manner of spoken input which will be acceptable to the system.

For isolated word recognisers a distinct gap (usually of about 300 ms) has to be left between each word, this is unnatural and requires concentration for commands which consist of more than one word.

For a connected word recogniser (Ref: 1) the words can be spoken normally i.e. without enforced gaps, provided that the phrases are kept reasonably short (usually less than 10 words). This is not expected to cause problems in an aircraft application as long command strings will not be used. It should be noted that although a connected word recogniser does not require gaps between the words, the speech must not be too rapid (as in normal fluent speech) because this causes the words to overlap and modifies the sounds at the word boundaries. A continuous speech recogniser can handle normal conversational speech i.e. unconstrained sentence length and the inclusion of words not in its vocabulary.

## 3.2 The Architecture of a Speech Recogniser for Aircraft Control

A typical speech recogniser consists of several discrete sections (Fig. 2), each one intended to convert its input into a form more related to the meaning of the utterance and less to the sound.

### 3.2.1 Preprocessor

This is usually a hardware device which converts the microphone signal into a form suitable for processing. To minimise the data rate, whilst preserving as much speech relevant information as possible, an encoder based on assumptions about the production or perception of speech is usually used, to discard non recognition relevant data.

### 3.2.2 Word Matcher

At present all practical speech recognisers use whole word pattern matchers, usually with some form of dynamic programming algorithm (Ref: 5) to account for variations in word length. The data for the word to be recognised is compared with a stored sample (template) of each word in the vocabulary and a score indicating quality of match is generated for each template.

As this presents a very large computational task, and puts equal importance on all input data whether or not it is relevant to the meaning of the words, two additional stages of processing have been tried; Segmentation and Feature Extraction. These have the potential of reducing the computation required, but have so far had only limited success due to a lack of knowledge of the nature of speech.

### 3.2.3 Word Spotter

In the simpler systems the word with the best matching score is chosen, in more advanced systems the words are selected on the basis of the matching score plus syntax and pragmatics (ie. data from the application).

Syntactical information can also be used to limit the size of the vocabulary searched by the recogniser at any given point in the dialogue.

### 3.2.4 Command Interpreter

The function of the command interpreter is to convert the string of recognised words into a command signal, in the format required by the aircraft systems. This would be performed by a look up table in current systems, but in an advanced system could be done using more intelligent software techniques.

### 3.2.5 Systems Interface

This is the hardware which connects the speech recognition computer to the aircraft systems under its control, and via which the recogniser receives syntactical and pragmatic data from the aircraft systems.

## 4.0 . APPLICATIONS

Initially, the limited active vocabulary typical of available recognition systems will constrain the command format to be similar to the controls it is replacing. An obvious and quite reasonable starting point is the CDU, since use of the keyboard requires significant periods of "hands off" "eyes in" operation. Although the command structure could be similar to that of the CDU, it should, where possible, be based instead on the present discrete systems operating procedure to give a better input sequence for speech.

If DVI is not made available for control of all the aircraft systems, the division should be clear and logical. It has been suggested that DVI could be used only for those items required during TTLF, but this division is difficult to define and would probably appear rather arbitrary at other times. Designing speech control to operate in parallel with the CDU should not produce problems, as the CDU itself will only be used for self contained systems with easily defined boundaries.

Command formats which are more related to the pilot's requirements than the aircraft equipment (see Section 6.4) can be developed when improved DVI systems become available. The scope for improved speech interaction will also increase when increased automation is incorporated in other aircraft systems and more information is available to the command interpreter, e.g. the waypoints for the required route could be described with reference to geographical features rather than a grid reference, if the navigation system included a digitally stored map.

## 5.0 DVI SYSTEM SPECIFICATION

There are several aspects of DVI systems not previously encountered in avionic systems, most of which concern the man-machine interface, and are thus difficult to quantify. This is particularly true of the recognition performance since the test signal is speech, which is hard to define and highly variable (discussed in "DVI System Testing," sect. 5.7).

### 5.1 Vocabulary Size

This is the maximum number of words which can be stored as templates. There may also be a smaller maximum number of words active in the recogniser at any one time, the others having been declared by the syntax system to be irrelevant to the context of the speech.

### 5.2 Isolated or Connected

The choice at present is between isolated and connected word recognisers. Although some isolated word recognisers can accept short phrases as a 'word', only connected word recognisers are considered practicable for aircraft use.

### 5.3 Recognition Control/User Interface

No standard control or interface exists, so each system must be considered individually. Factors to consider are the manner in which the syntax or grammar is programmed, the ability to use information from the application interface, and the format of the data which is available at the systems interface. This last aspect is particularly important for airborne applications since most systems do not include a command interpreter, and hence only supply a string of recognised words.

5.4 Recogniser Training Procedure

The users (i.e. the crew) for the duration of a flight are known, thus a speaker dependent system can be used. This does not necessarily imply a training period prior to flight, as each crew member could carry a personal voice data memory pack. Speaker dependent systems with acceptable vocabulary and reliability are available now and need retraining at intervals, so the training procedure must be simple and preferably automatic, using only one sample of each word.

5.5 Interfacing

The interfacing requirements of a DVI system are unusual in that it requires signals from most other aircraft systems, including those it is not controlling. This data requirement arises because the use of contextual information improves the recognition performance and can allow a more natural interaction with the DVI system.

5.6 Noise Immunity

Although the cabin noise level may be high, say up to 115dB, most of the noise is of a repetitive nature from the engines and transmission etc., which is readily removed, whereas in fixed wing aircraft more of the noise is of aerodynamic origin.

A typical noise spectrum, as measured on the pilot's microphone line, is shown in Fig. 3. This shows that apart from several discrete frequencies, which could easily be filtered out, there is an adequate signal to noise ratio throughout the relevant speech band.

5.7 DVI System Testing

Speech recognition system testing is a field in which little real progress has been made to date. The only method currently used, is to measure the percentage recognition performance using a standard data base of recorded speech. Various groups have carried out comparative tests on commercially available devices, but whilst the results give a reasonable guide to the relative performance of the equipment tested, the results only apply to the given vocabulary, speakers and conditions. Recognition figures given for these tests are usually between 86 and 99.5% (Ref: 6 & 8) for the various systems. This compares well with error rates of around 10% for touch operation of numeric keypads (Ref: 7). Performance figures in noise, particularly important for aircraft applications, are usually of very little use since often only the sound pressure (dBA) is quoted, a change in the microphone used and the particular type of noise can clearly cause widely different results. A standardised testing procedure is required for aircraft applications, both for the customer and the development teams, particularly with regard to noise conditions. The only nearly satisfactory test is one devised for the final application, using a realistic vocabulary and representative noise. An acceptable error rate can only be established in an accurate simulation of the application but may be as high as 2%.

A significant problem with the use of data bases is that if a system is modified on the results of such a test it must then be regarded as optimised for that data base, and thus it would no longer be a fully objective data base if used repeatedly. Work is at present underway on guidelines for testing with data bases (IEEE), and on alternative methods (Royal Signals and Radar Establishment, Malvern, England).

## 6.0 HUMAN FACTORS

DVI is a man-machine interface and the human factors advantages i.e. redistribution of workload (see 2.3), are the principal reasons for its use and clearly of prime importance in the design. It is the human factors and the application which define the syntax structure, the command interpreter and the application interface. Thus a thorough understanding of the human factors aspect is required before the system specification can be written.

The development of an effective operating protocol will require the operation of a test system, either in the real situation or preferably in a realistic simulation. Since one of the main advantages of DVI is 'eyes out' operation this implies experimental test and validation on a flight simulator with a good vision system.

Some of the human factor aspects under investigation at WHL are;

### 6.1 The Correct Form of Feedback

The user will require feedback to confirm that his commands are being input. The form often suggested in this case is speech output, which is technically feasible, but not necessarily ideal. Some pilots comment that direct voice output (DVO) is an improvement over orthodox audio warnings and advisories, but the excessive use of DVO could interfere with voice comms monitoring. The impermanence of DVO information should not cause problems since the relevant information will also be displayed by the CDU or CWP.

However use of a head down visual channel for feedback removes a major advantage of DVI, in order to alleviate this one option is to install a display near the sight line. The options for this display are;

- o on the coaming
- o on a head up display
- o on a helmet mounted display,

The last two are probably only worth considering if the display would also have other uses, as the cost and weight penalties could probably not be justified for this application alone.

6.2 Voice Variation

Voice variation on a day to day basis, e.g. due to ill-health, has not so far proved a serious problem, particularly with experienced users, but variations which might occur during a flight i.e. due to vibration, stress etc. are not yet particularly well explored.

6.3 Speaker Co-operation

Pilots' experience in the use of noisy communications channels gives them a discipline in controlled speech, which should prove of benefit to DVI Systems.

6.4 Command Structure

The DVI command structure could be designed to accept exactly the same command phrases the pilot would use when instructing the co-pilot. The pilot would then not need to remember an arbitrary input sequence, or refer to a prompting display.

The development of an effective command structure is dependent on:

- o The ability of the recogniser to handle a wide branching syntax. This will allow the use of shorter command strings but will put a higher load on the recogniser by increasing the active vocabulary. This also increases the computational load and the probability of confusing two words.

- o The development of sophisticated command interpreters and systems interfaces (section 3.2) that will translate commands that are relevent to the pilot's requirements (i.e. "connect me to BLUE One"), into signals compatible with the aircraft equipment (i.e. "VHF 1 set frequency 165.75, set mode AM, secure").

## 7.0 PERFORMANCE BENEFITS

The main advantages of a well designed DVI system will be a reduction of the visual and manual workload, and improved interaction with the aircraft systems. The decrease in visual workload will give increased observation time, this should lead to improved survivability. The ability to make more effective use of the avionics systems without compromising the observation, piloting, and weapons operation, will give increased operational effectiveness. With the increased automation of aircraft systems and the reduction in manual workload given by DVI, single crew operation should become possible in an increasing number of roles.

## 8.0 CONCLUSIONS

The performance of direct voice input systems has reached a level to enable serious aircraft applications research and development to be undertaken. The developments in signal processing techniques together with continued improvements in hardware will result in cost-effective airborne systems being available in the next three to five years.

To fully exploit the benefits of DVI in the helicopter environment, the voice control system commands need to be integrated within a total avionic command and control 'data base' to provide the necessary contextual information and to allow greater freedom in commands.

System test procedures are seen to be a major area of work which needs to be addressed in the short term. Test techniques need to be developed which are compatible with the variations in speech from the population of pilots and aircrew likely to use a system over the life of the system and aircraft.

## 9.0 REFERENCES


1. Bridle, J. S. and Brown, M. D., "Connected Word Recognition using whole Word Templates". Proceedings - The Institute of Acoustics, Autumn 1979.

2. Coler, C. R., "Helicopter Speech-Command Systems: Recent Noise Tests are Encouraging". Speech Technology, Vol. 1. No. 3. (Oct. 1982) p.p. 76 - 81.

3. Sexton, A. S. and Jones, T. G., "Voice Command and Response: new Solution to New Problems", ICAO Bulletin, March 1981.

4. Velichko, V. M., "Automatic Recognition of 200 Words", Int. Man-Machine Studies 2 222-234, 1970.

5. Sakoe, M., "Automatic Continuous Speech Recognition Employing Dynamic Programming", U.S. Patent 4,057,725, 1976.

6. Moore, R. K., "Evaluating Speech Recognisers", IEEE Transactions on Acoustics, Speech and Signal Processing 25, No. 2 (1977) p.p. 178 - 183.

7. Nakatani, L. H. and O'Connor, K. D., "Speech Feedback for Touch Keying", ERGONOMICS, 1980 Vol. 23. No. 7, p.p 643 - 654.

8. Baker, J. M. and Bridle, J. S., "Speech Recognition Performance Assessments and Available Data Bases", Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing, 1983, 12.2 p.p. 527

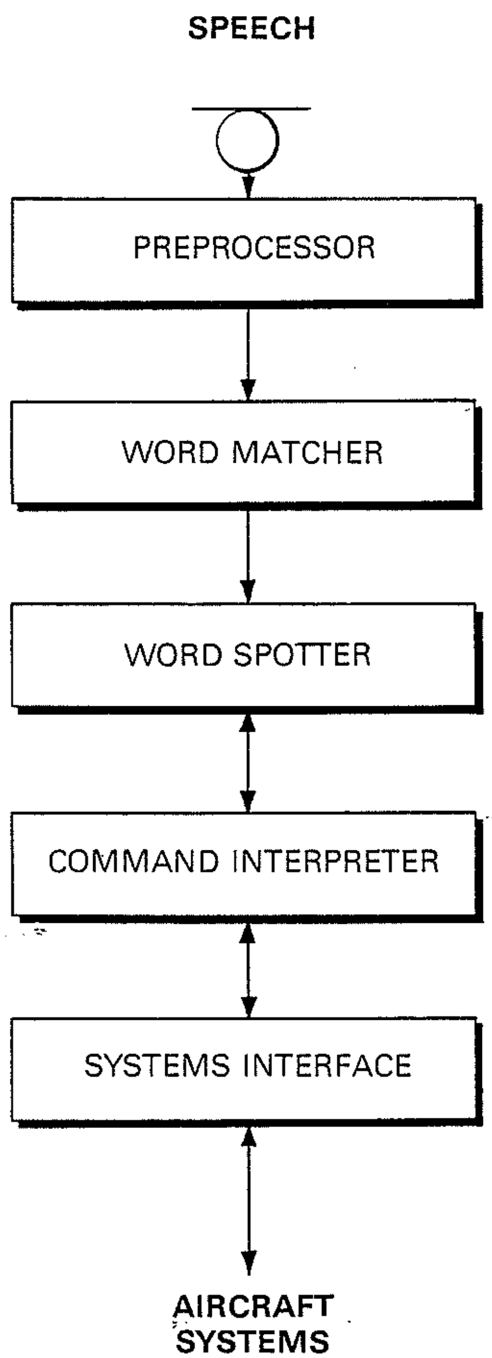


FIG 1
DVI SYSTEM COMPONENTS

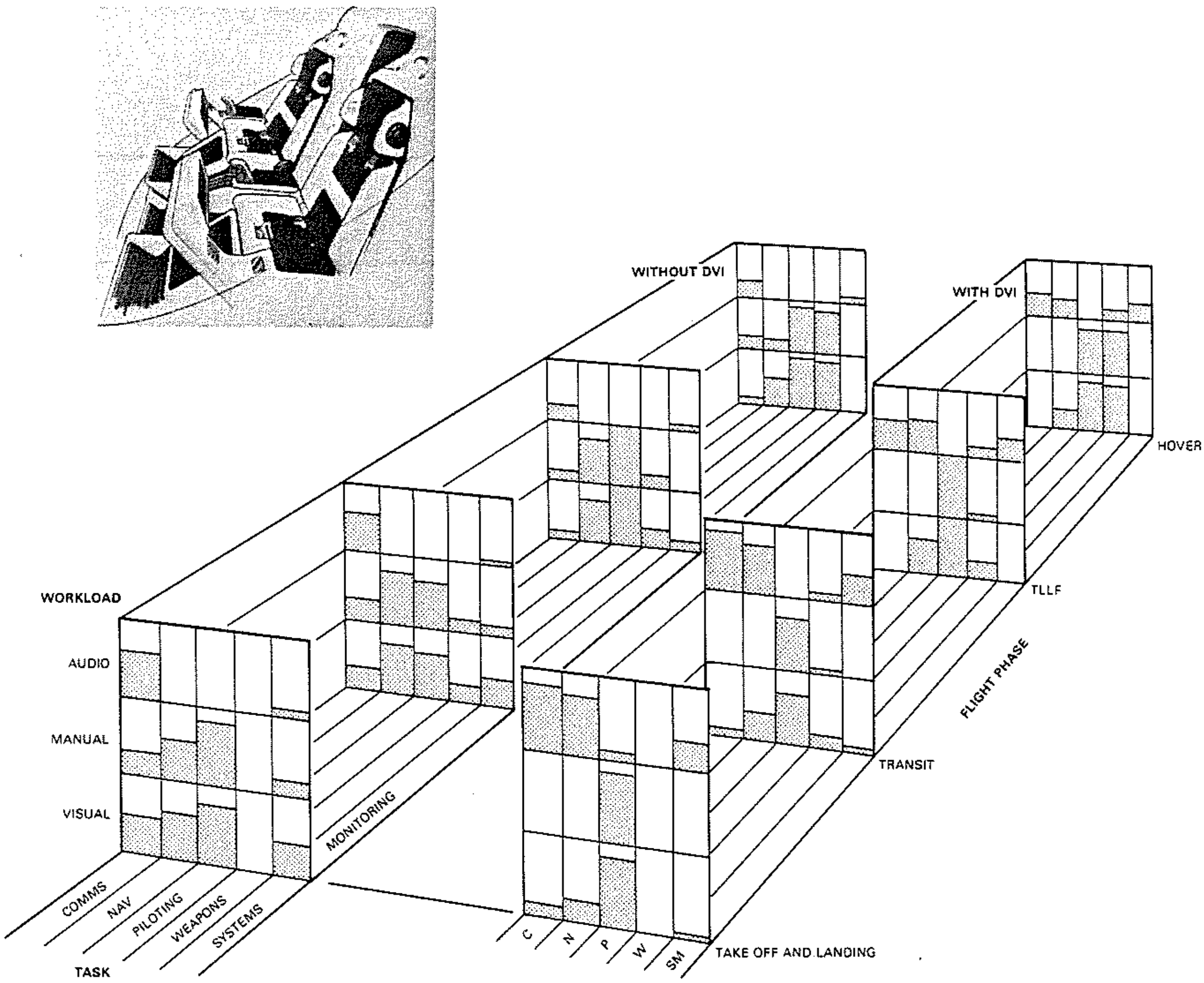


**FIGURE 2**
**CHANGE IN WORKLOAD WITH THE USE OF DVI**

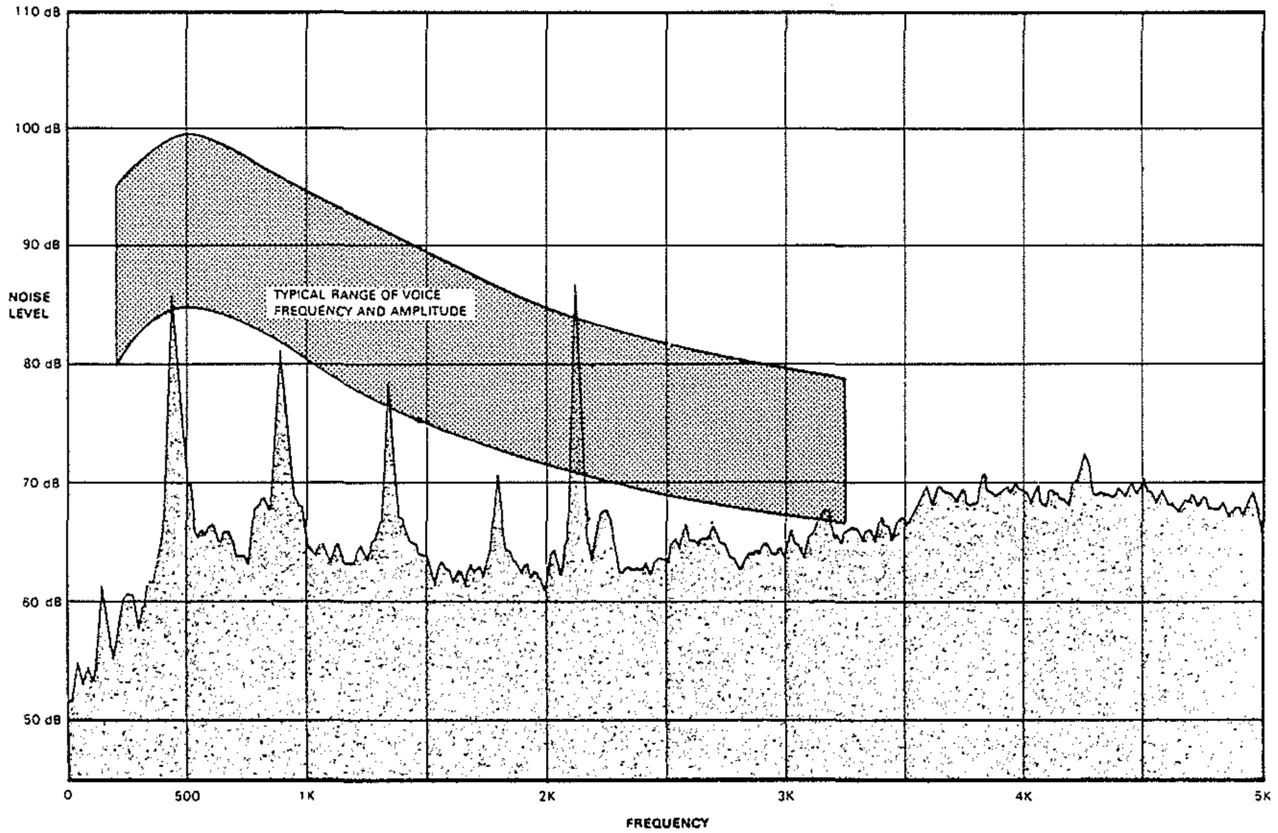


**FIGURE 3**
**NOISE IN LYNX COCKPIT, LEVEL FLIGHT THROUGH THE HEADSET BOOM MIC**