

# DEVELOPMENT CONSIDERATIONS FOR SPEECH BASED HEARING TEST MATERIALS FOR FLIGHT CREW

J.E. Clark, C. Whitfield, and P.J. Kennedy

Speech Hearing & Language Research Centre

Macquarie University

**ABSTRACT** - Traditional hearing tests for technical flight crew have relied on the conventional pure tone threshold tests commonly used in clinical audiometry. This paper describes some of the considerations involved in the development of a set of speech based hearing test materials designed to evaluate the capacity of flight crew to satisfactorily process the speech signals routinely used under operational conditions in aviation. The rationale for the development of such tests is the recognition of the limitations in pure tone tests for distinguishing between crew with functionally effective hearing, and those who are no longer functionally effective when their pure tone thresholds are around the borderline of the ICAO limits. Factors investigated in the process of test development are described, including communication systems properties, cockpit noise levels, and operational language characteristics. The need for three classes of test is shown, and some of the considerations and procedures for their compilation are described.

## INTRODUCTION

This paper reports on aspects of the development of a speech based test for use with technical flight crew in the Australian context. It has its origins in concerns that the pure tone threshold tests as prescribed in the relevant ICAO standard (not more than 35 at any one of the three frequencies 500, 1000, and 2000 Hz, and not more than 50 dB at 3000 Hz) may not provide a realistic assessment of the functional communicative capability for flight crew whose thresholds are at the margin of the standard. It is not the purpose of the present test development to supplant the pure tone based standard test, but rather to supplement this with a functional test of communicative competence based on a simulation of actual listening tasks which are pertinent to air safety.

The perceived deficiencies of the pure tone tests include;

- no account being taken of communication system characteristics and limitations;
- no account of complex the auditory processing workload of flight crew under operational conditions;

Historically, some speech testing has been used for flight crew in Australia, and is described in Dermody & Mackie (1987). This involved utilising existing PB word lists in conjunction with masking noise designed, amongst other things, to simulate DC-3 aircraft propeller and engine noise. Both Dermody & Mackie and Carter & Farrant (1957) have pointed out the limitations of these materials. Presentation levels were either at 108dB SPL (cockpit simulation) or 60 & 70 dB (office briefing simulation). Perceived limitations of these early speech-based tests include:

- no account being taken of the wide range of operational conditions in modern aviation;
- inappropriate noise characteristics;

- inappropriate language materials given the known restrictions on operation vocabulary and syntax.

In-flight testing has also been used for a number of years, but has very obvious disadvantages in terms of cost, convenience and reliability.

The test materials developed in the present study were designed to overcome a number of these difficulties, and to provide a valid simulation of the general range of conditions under which operational listening tasks are currently undertaken in the Australian aviation industry.

To establish the simulation, factors examined included; task-specific vocabulary and syntax, transmission path electro-acoustic properties, system and propagation path noise, and environmental acoustics, and (in varying degrees) accounted for. Substantial background material on these factors which were subsequently utilised in test development may be found in Koob & Clark (1986) and in the language and environmental acoustic analyses of operational procedures and conditions in Kennedy & Clark (1988a, 1988b).

## TEST CLIENTELLE

Following discussions with the Dept. of Aviation it was agreed that the tests be designed for personnel described under Air Navigation Regulations 51 and 52, which concern personnel holding licences to engage in commercial aviation operations. These are:

- (Senior) Commercial Pilots
- Airline Transport Pilots (Classes 1 & 2)
- Flight Engineers

## CLASSES OF TEST REQUIRED

One major motive for the studies of cockpit noise and operational language described in Kennedy and Clark (1988a, 1988b) was to ascertain the number of tests required to meet the range of operational conditions found in the aviation industry. There is, of course, no single optimal answer to that question. Cockpit noise levels alone vary widely even within individual aircraft types, depending on current operation status (e.g. take-off, climb, cruise etc.) However examination of data collected on 44 aircraft indicated that it would not be unreasonable to establish three relatively natural classes of test which would reflect both logical operational groupings and substantially different speech communication acoustic environments. These were:

- Class 1 - General Aviation (single and twin engine piston or turbo-prop monoplanes)
- Class 2 - Regular Passenger Transport (tail fuselage and wing mounted turbo-jets)
- Class 3 - Rotary wing (piston and turbo-engine)

## ESTABLISHING TEST AND ASSOCIATED SIMULATION CHARACTERISTICS

### Class Average Cockpit Noise Data Noise

For each of the above classes the cockpit noise spectra were chosen to provide a presrepresentative approximation of the entire aviation class, and the more difficult listening conditions encountered during normal operations. Consideration of the duration and significance of operationsla conditions was also taken into account, with the result that taxiing, takeoff and landing conditions were not utilised. Relatively low altitude climb and descent conditions proved the most suitable, both in terms of high noise level and

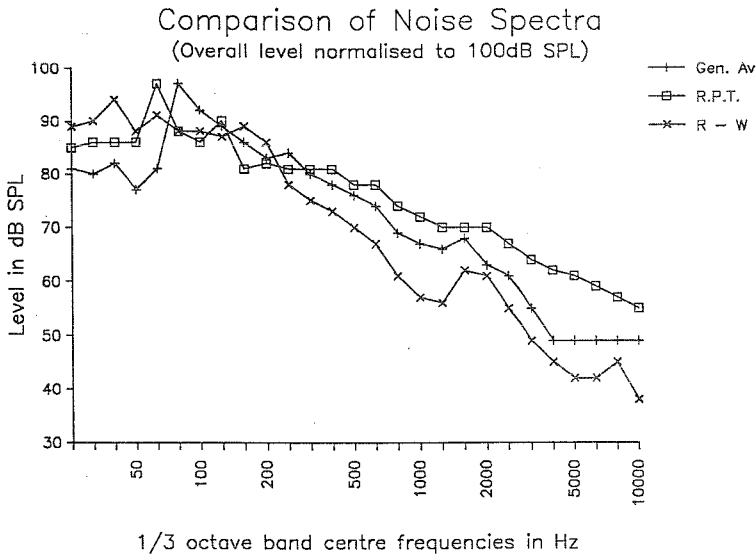


Figure 1: Comparative long term noise spectra

high significance to aviation safety, as they are commonly associated with approach/departures to/from controlled airspace. Figure 1 shows the average long term noise spectra for the three aircraft classes in the test categories chosen.

Despite the apparent similarity of these spectra, their dynamic properties are quite different. The RPT noise is largely aperiodic resulting from air turbulence; that of general aviation piston engine aircraft has substantial low frequency (80-100Hz) propeller noise with an appreciable harmonic component together with turbulence; and rotary wing aircraft noise is dominated by rotor noise in a mid band range (around 600Hz). As well as 1/3 octave band analysis, spectrographic analyses were conducted on selected samples to ensure that there were no major gaps in spectral energy distribution.

The production of appropriate simulations for the noise masking in the three test classes were explored using:

- digital synthesis using white noise shaped by a defined spectral envelope;
- analogue synthesis using 1/3 octave band filtered pink noise;
- remixing and filtering in-flight recordings of noise.

Each of these approaches had difficulties. The first two, while able to replicate long term aviation class spectra, lacked the necessary harmonic components, and hence realism. The last method showed marked inconsistencies of level even with multitrack recordings (especially in the general aviation class) and would not have provided predictable masking levels.

The eventual choice was the use of analogue signal generation based on a pink noise source mixed with a triad of pure tones from independent signal generators. This ensured random phase relationships of the kind found in complex cockpit noise arising (as is commonly the case) from several disparate sources.

A final consideration relating to cockpit noise concerned the attenuation effects of headsets worn by crew. Observation of actual headset use showed that in general aviation there was no consistent wearing of headsets. RPT crew often used "custom" headsets with little attenuation and in many cases only covered

one ear. It was therefore decided that no allowance would be made for headset attenuation effects in these test classes. Rotary wing operations, in contrast, showed universal headset use and an increasing use of full helmets incorporating headsets. The use of a headset alone was assumed the worst case, and the class test noise conditions based upon that premise. Data supplied by the National Acoustics Laboratory on the attenuation properties of a representative headset commonly worn by rotary wing crew was used to correct the basic average class cockpit noise data for the purposes of simulating the masking in this class.

### Language Materials

The analysis of operational language reported by Kennedy and Clark (1988a) reported that 382 lexical items accounted for 95% of all words used in an analysis of 19,632 words from 1,726 messages. Further analysis showed that of the 382 items, some 222 were common to the three proposed test classes. It was therefore decided to examine the use of a common set of language materials as potential test items for the three classes.

Based on the Kennedy & Clark analysis referred to above, the following materials were compiled:

- Isolated words;
- a three word sequence (e.g. alpha charlie delta)
- an aircraft callsign followed by an operational message up to five words (e.g. alpha victor romeo clear for takeoff)
- an aircraft callsign, a two to four digit sequence and an associated carrier phrase (e.g. foxtrof india zulu *make your speed two six zero knots*, where the words in italics form the effective carrier phrase).

The isolated words are intended to test basic speech discrimination ability with minimal linguistic context, and the messages are designed to assess more complex auditory information processing capacity including short term memory.

### Speech Filtering

The R/T transmissions systems used in aviation have a relatively narrow bandwidth and an analysis of the long term spectrum of radio traffic signals allowed this to be simulated using a 1/3 octave band filter set. Unlike air traffic controllers, aircrew use a wide variety of headsets with quite differing frequency responses, and it was judged better not to attempt to account for this as there was as much variability with as between test classes. Moreover, the major constraint on signal bandwidth appeared to be determined by the radio transmission system rather than subsequent acoustic transduction.

### Radio Transmission Noise Simulation and Signal to Noise Ratio

Samples of HF and VHF noise from the three test classes showed sufficient similarity within the passband of interest to enable the use of a single shaped spectrum for all tests. This, of course is a separate issue to the ambient cockpit noise properties described earlier.

Ascertaining actual signal to noise ratios is not straightforward because of the rapid action automatic gain control circuits employed in receivers. However, use of a digital speech editor system allowed the analysis of short signal sequences from which signal to noise ratios could then be calculated. As expected, these ratios varied widely. Initial figures of 0dB, +3dB and +6dB were used to trial test materials, and represented approaching worst case conditions.

### Preferred Listening Levels

Preferred listening levels were investigated for each test class. Crew rarely reported inadequate gain adjustment availability in actual operating conditions, so that wide flexibility was allowed in the investigation. Some 65 aircrew were tested across the three classes. An ATIS (automatic terminal information service message) was used as the narrow band speech signal source and an up-down procedure

employed in which aircrew adjusted a stepped precision attenuator system to set a preferred listening level. The mean preferred listening levels were:

- 101dB SPL with a simulated cockpit noise level of 106dB SPL for the general aviation class;
- 85dB SPL with a simulated cockpit noise level of 90dB SPL for the RPT class;
- 86dB SPL with a simulated cockpit noise level (including headset attenuation correction) of 99dB SPL for the rotary wing class.

## **COMPILING TEST MATERIALS**

### **Speech Material Recording**

The test material was spoken by both male and female talkers. Experienced air traffic controllers were used to ensure maximum realism and the intuitive use of speech prosody patterns normal to aviation operations. Both were selected to have the following attributes:

- general Australian accents;
- normal articulation patterns
- normal rate of utterance.

The controllers also cross-checked all the language materials to ensure they reflected current air traffic control practice and to ensure there were no anomalous items. The basic recordings were done in a sound treated recording studio using a Betamax PCM system. All items were checked and re-recorded where any aberrant articulation occurred. Long term spectral analyses of the talkers recordings were also made to ensure they conformed with more general data obtained in the SHLRC and with the classic data of Dunn & White (1940).

### **Speech Material Processing and Transduction**

The speech test materials were processed and compiled using special purpose speech signal editing software. The initial speech material was segmented to remove unwanted sequences and was also level normalised, three files of data were then employed in the actual test materials. These were; a speech file containing the segmented and identified speech signals, an RT noise simulation, and the three cockpit noise class simulations. Final output files of test material contained a 12 second 1KHz calibration tone, 5 seconds of lead-in cockpit noise, and lists of test items at varying signal to noise ratios. These latter were finally determined by the listening test trial results and could be varied on an item by item basis.

When considering the electro-acoustic transduction of the complex speech and noise signals involved in this test, it was deemed impractical to establish realistic free field sound replications of cockpit noise for clinical purposes. Moreover, it became apparent that standard audiometric headphones such as the TDH-49 with MX41/AR cushions lacked the low frequency response needed for cockpit noise simulations. Beyer DT48 headphones are included in Australian Standard AS 1591 and had the necessary extended low frequency response when used with circumaural cushions to ensure both fidelity and accuracy in setting environmental SPL values. They, inter alia, the German Standard (DIN) reference headphone, and have a frequency range of 16-20,000Hz.

## **TEST DEVELOPMENT**

Details of test development, trialling and results will be the subject of a further paper and cannot be reported in detail within the scope of this paper. Briefly, the objective has been to produce a test ensemble of maximum sensitivity by iterative adjustment of items to produce a similar degree of difficulty. The rationale for this is described by Dillon (1983).

In setting test conditions and adjusting item difficulty, the available variables include:

- stimulus test level;
- ambient (cockpit) noise level
- transmission signal to noise ratio.

Given that the first two are determined by simulation fidelity considerations, the third variable provides the appropriate control mechanism for item difficulty. The general procedure followed was to establish item difficulty characteristics using a trained test crew of university students with normal hearing. Those items with appropriate intelligibility characteristics were then used in the compilation of the test materials proper. It should be noted that some test items have inherent perceptual salience properties which make it difficult or impossible to set appropriate difficulty levels without using inappropriate (i.e. unrepresentative) signal to noise ratios.

The characteristics of a select group of test items were then evaluated using flight crew, and a final set of test materials compiled following appropriate item difficulty adjustment to approximate the Dillon criteria. In all cases, an equal number of male and female voice test items were utilised.

## CONCLUSION

A set of considerations used in the design of special functional hearing test materials has been described. This process has involved careful scrutiny of all aspects of the speech communication process within the context of the operational environment found in commercial aviation. It is expected that the resultant tests when clinically validated will provide a better general representation of actual listening task conditions than has been available previously.

## REFERENCES

- Carter, N.L. & Farrant, R.H. (1957) Australian replication of ICAO aircrew hearing standards. Commonwealth Acoustics Laboratory Report No. 12. Sydney.
- Dermody, P. & Mackie, K. (1987) Speech tests in audiological assessment at the National Acoustics Laboratories. Ch. 14 in Martin, M. (ed.) Speech Audiometry, Taylor & Francis, London.
- Dillon, H.K. (1983) The effect of test difficulty on the sensitivity of speech discrimination tests. Journal of the Acoustical Society of America, 73, 336-344.
- Dunn, H.K. & White, S.D. (1940) Statistical measurements on conversational speech. Journal of the Acoustical Society of America, 11, 278-288.
- Kennedy, P.J. & Clark, J.E. (1988a) Operational language in the cockpit/flightdeck communication environment of Australian civil aviation aircraft. Proceedings of SST-88, 268-272.
- Kennedy, P.J. & Clark, J.E. (1988b) Ambient noise in the cockpit/flightdeck communication environment of Australian civil aviation aircraft. Proceedings of SST-88, 274-279.
- Koob, P.C. & Clark, J.E. (1986) An investigation into the formulation of a speech intelligibility test using item difficulty criteria. Proceedings of SST-86, 150-155.