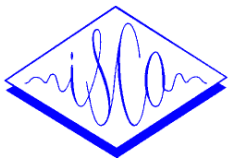

ESCA-NATO Workshop on
Speech under Stress
Lisbon, Portugal
September 14-15, 1995

Automatic Speech Recognition Systems: effects of environmental stressors

Chris Baber and Jan Noyes†

School of Man & Mech Eng, University of Birmingham,
† Department of Psychology, University of Bristol

ABSTRACT

There are a number of types of environmental stressor which can have an adverse impact on the use of automatic speech recognition. In this paper, a systems approach to the problem is presented, in which the system comprises the human and the speech recognition device performing specific tasks in a defined environment. From this perspective, the locus of the effects of environmental stressors can be seen to be the interface between human and speech recognition device. This means that it is important to appreciate the effects of the stressors on both human performance and speech production, and the likely consequences of these effects for the efficiency of the speech recognition system.

INTRODUCTION

'The task environment comprises a number of factors that must be studied for their effect on human performance and therefore on speech task design. Physical, physiological, emotional and workload factors can be expected to partially determine the success or failure of a particular speech system.'

Simpson et al. [1]

It is proposed that adverse environments represent a distinct class of application domains, and that many of the problems require detailed consideration of human factors solutions, in addition to technological solution [Baber and Noyes, 2]. Characteristics of different domains, in industry and military applications, will be related to proposed human performance effects and various solutions will be presented for consideration. The aim of the work is to raise the level of awareness of the potential problems related to human performance effects of 'stress' and automatic speech recognitoin (ASR).

It is clear that the characteristics of some types of stress can have an impact on the production of speech. For instance, acceleration and vibration will impose changes to physiology of speech production (for instance, leading to shakiness in the voice), while noise will have implications for the monitoring of speech (leading to an increase in speech intensity). However, in order to appreciate the effects of the stressors on the human-computer system, it is important to understand the effects of stress on human performance.

In situations of severe levels of environmental stress, human performance degrades such that the performer tends to focus on a limited range of cues, uses familiar (rather than appropriate) responses, is prone to error when following procedures and has disrupted short term memory. The implications of such effects, for speech recognition systems, are that users may forget or confuse appropriate command words, may not be able to recall or follow the appropriate command syntax, may not be able to effectively monitor recognition feedback, amongst others.

Driskell and Salas [3] provide a detailed review of the effects that environmental stressors have on human performance. Table one lists some of the principal physiological, physical, affective and psychological effects induced by adverse environmental factors:

Physiological -	*Physical* -
•increased heart rate; •laboured breathing;	•trembling; •reduced manual dexterity; •muscular contraction;
Affective -	***Psychological*** -
•loss of motivation; •inability to act;	•reduced efficiency in search behaviour; •restriction in number of cues selected; •increased reaction time; •reduction in vigilance performance; •reduction in problem solving ability; •performance rigidity; •increased errors in following procedures; •disruption to recall from working memory.

Table one: Effects of stress on performance

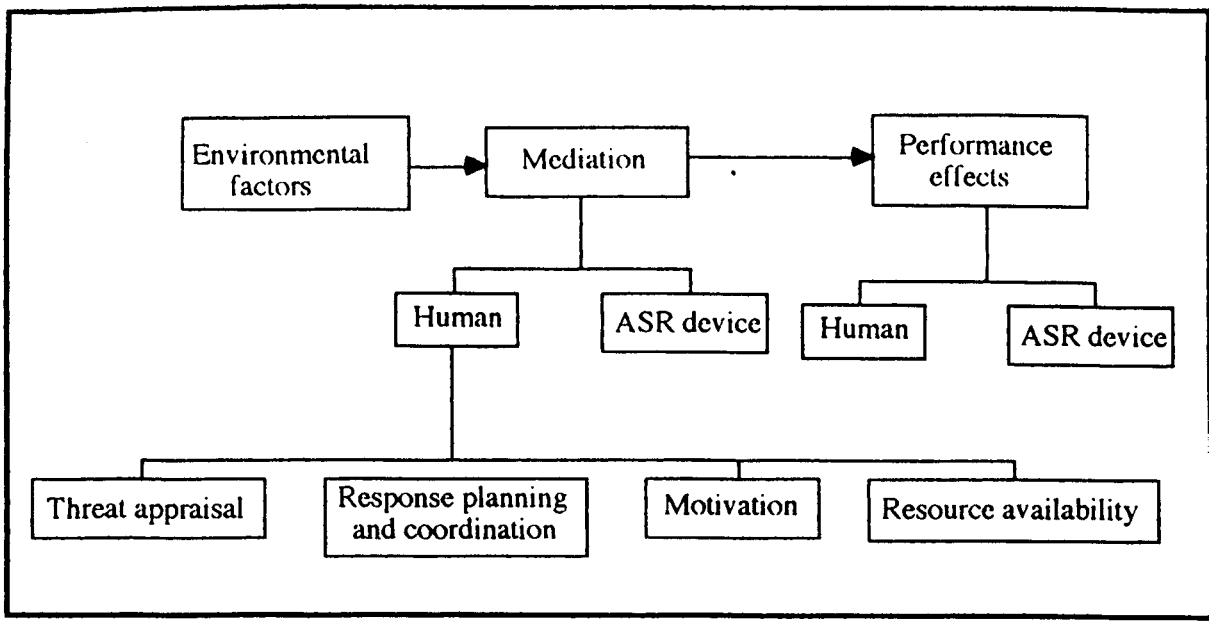

Figure one illustrates the relationship between environmental factors and the concept of mediation (Baber and Noyes, 2). Figure one suggests that mediation can be performed by the human (Noyes and Baber, 4) or by ASR devices, e.g., in terms of noise cancelling or other means of reducing contamination of incoming signals from environmental factors. There has been a great deal of research effort aimed at implementing such device mediation, particularly with respect to ambient noise, and efforts have been directed at building ASR devices which are robust enough to cope with the normal cockpit environment, e.g., to be able to withstand extremes of vibration, pressure, acceleration etc. Several ASR companies are now marketing devices which have been tailored for specific application environments, such as helicopters or high performance aircraft. Furthermore, there has also been some effort aimed at reducing the direct effects of environmental factors, such as noise and acceleration, on human performance with ASR. However, it is apparent that there has been relatively little attention given to the potential effects of indirect environmental factors, such as stress and workload, on ASR use (Baber, 5). The key points to note are that environmental factors will have effects on the performance on both the human and the speech recognition device, and that these effects can be modified the behaviour of the person and by the recognition algorithms of the device. In the remaining sections, the effects of acceleration, vibration and noise are considered.

ACCELERATION

Under very fast acceleration, i.e., in excess of + 6 Gz., speech articulators in the vocal tract can be displaced. As one might expect the effects of acceleration on speech production are symptomatic of increases in forces applied to the speech production apparatus and to changes in breathing rates. Displacement of the vocal tract as a result of increased acceleration leads to marked shifts in the formant frequencies of speech. The shift resulting from acceleration is likely to be exacerbated by the effects of wearing an oxygen mask, although it is not clear how stressors interact.

There are recognised measures for coping with high Gz., such as adopting specific postures or wearing special clothing. However, these measures can also have an influence on the pilot's ability to produce intelligible, consistent speech. As far as ASR is concerned, it would appear that the principal effects of acceleration will be physiological and probably beyond human control. Thus, it is necessary to ensure robust template matching, either through training pilots in effective ways of speaking at high Gz., or enrolling 'operational speech' under simulated acceleration conditions, or through well designed matching algorithms. It might be possible to consider the range of speech functions required under acceleration and to produce a minimally confusable template set solely for such functions, although given the complexity of cockpit management tasks this is probably easier said

than done. This leaves the principal solution to recognition problems lying with the developers of ASR equipment. Several ASR manufacturers are developing equipment which can withstand extremes of environmental operation, although it is a moot point as to whether this development is aimed solely at the ASR device or at the human-ASR system.

One might object to the use of ASR on the grounds that acceleration disrupts speech, but performance using manual input media will also be affected, possibly at lower levels of Gz. Thus, ASR could potentially provide for faster response than manual input. We are not aware of any research which compares the efficiency of task performance using ASR and manual input devices under conditions of high Gz., so this issue remains a moot point.

Given the potential human performance problems outlined in table one, one can envisage detrimental effects on ASR relating not only to the physical production of speech, but also to such aspects of ASR use as feedback monitoring. For instance, several applications of ASR in cockpits rely on visual feedback to present ASR response to pilots. Given that acceleration has an influence on some aspects of visual perception, it is conceivable that pilots will misread feedback, especially when one considers that visual feedback is prone to misreading. Given also the possibility of disruption to immediate memory caused by high Gz., it is likely that care should be taken in the provision of adequate feedback for ASR operation under such conditions, i.e., feedback which is not only easy to read but also displayed for sufficient time for the pilot not to have to rely on short term memory.

VIBRATION

Considering the potential effects vibration will have on the production of speech, one can assume that variations in airflow through the larynx will probably lead to some degree of 'warbling' in speech sounds, and that some frequencies of vibration will induce breathing irregularities which, in turn, will impact on speech sounds. Further, general body tension resulting from an individual's effort to deal with vibration coupled with interference to the movement of the lower jaw can also be proposed to influence speech production. Having said this, studies conducted at relatively low levels of vibration in laboratories suggest that vibration need not be a problem. However, there appears to be some contradiction in studies conducted at operational levels. For instance, the level of vibration encountered in fixed wing aircraft lies between 10 Hz. and 22 Hz., and these levels are sufficient to interfere with speech production.

Under low level, high speed flight vibration would be sufficient to buffet the body and lead to changes in speech pattern, i.e., fundamental frequency increases and the space between F1 and F2 becomes more compact . The effects will be beyond human control, so difficulty will be encountered in maintaining consistent speech.

Given that it will be difficult to maintain consistent speech patterns under vibration, the question arises as to what measures ought to be taken. Clearly the practice of enrolling speech under simulated vibration is possible, but different levels of vibration can produce quite different effects on a person's speech. Thus, defining a simple, standard vibration level of enrolment would appear to be impractical. Alternatively, as mentioned in connection with acceleration, it might be possible to develop some form of adaptation to cope with the fluctuations in speech produced under vibration, although such research has yet to be conducted.

Due to the limited amount of research specifically related to the effects of vibration on the use of ASR in operational environments, it is difficult to point to other human performance problems. There is an implication that the ability to verbalise responses could be impaired, which obviously has serious implications for ASR, and that visual monitoring of feedback could be a problem a some levels of vibration, possibly to a greater extent than with acceleration.

NOISE

It is fair to say that noise is the environmental factor which has received the greatest degree of attention from the ASR community, both from human factors specialists and from engineers. This is probably due to the fact that noise is a characteristic of many of the application domains for which ASR has been considered.

Noise levels in excess of 80 dB(SPL) can lead to increases in the amplitude, duration and vocal pitch of speech relative to quiet, with changes also observed in formant frequency. Bearing in mind that the typical noise level encountered in a UH-60 Blackhawk helicopter is in the region of 103 dB(A), it is clear that noise is a non-trivial problem for speech applications.

A point to note is that many of the solutions for dealing with noise appear to share the assumption that noise is, to some extent, stationary, i.e., noise which is continuous and occurs within definable limits. Given this characteristic, it ought to be possible to provide a simple means of compensating for noise. There is ample evidence to indicate that enrolling an ASR device in the same ambient noise levels to those in which it will be used will reduce the possibility of recognition error, and for a number of writers this is sufficient to establish this point in guidelines.

However, there are many environments in which noise is non-stationary, e.g., the cockpit of a fighter aircraft. In these instances, additional measures will be required. These measures can either be applied prior to processing or post processing.

Studies in the mid 1980s focussed on the possibility of reducing noise effects prior to processing. A common approach was to aim for cancellation of noise effects, either in terms of active noise compensation or noise masking

There is an additional problem which may not be so easy to solve; in noise, people tend to increase the volume of their speech in order to hear what they are saying. This is known as the Lombard Effect, and is dealt with elsewhere in the proceedings.

At this point, it is worth noting some of the literature relating the effects of noise to human performance. Contemporary research suggests that people deal with noise through the adoption of specific task performance strategies,s and appropriate strategies can be developed which act as buffers to the detrimental effects of noise.

Taken together these findings suggest that noise could impair the use of ASR in situations where the vocabulary was perceived to have a low level of salience to the task in hand, i.e., users may attempt to use words which they consider more salient, and in situations where speech is not compatible with the task in hand, e.g., using speech for spatial orientation. This means that noise is likely to exacerbate the problems of poor system design.

The effects of noise on ASR use have been investigated by a number of researchers, and their efforts are now beginning to come to fruition. Commercially available ASR devices are being equipped with sophisticated noise cancelling and adaptation algorithms, which minimises the effects on noise on performance. However, there remain some potential human performance effects which noise may induce, principally either physical or psychological, and suggest that these factors will also require consideration in implementing ASR.

DISCUSSION

This paper inevitably raises the question of when should ASR be used. Although this point lies outside the discussion in this paper, it is worth noting the many cockpit based ASR systems are used for such 'house-keeping' tasks as radio and navigation frequency selection. It is feasible that such activity is most critical at low level, high speed flight, especially when the pilot must fly nap of the earth. The speed with which the aircraft moves between sectors means that communications need to be constantly updated, the need for head-up flight in this situation suggests speech as an ideal medium for control. This discussion suggests that there is still much research needed before we can emphatically endorse ASR in such a context.

The perspective taken in this paper is that the application of ASR needs to take into account the effects of environmental stress on the speech recognition system, i.e., on the interaction between human and speech recognition device. By exploring some of the potential effects on the performance of human and device, it is possible to develop strategies by which these effects can be eliminated or mediated.

Given that mediation is a potential approach to dealing with variation in human performance, and that it is possible for appropriate strategies to be defined, it becomes important to consider the interaction between individual differences and stressors in ASR systems. This would require a move beyond the conventional 'sheep' / 'goat' dichotomy of ASR research, towards a more detailed model of how stressors are perceived and mediated by individuals varying in terms of training, motivation etc. The design of the total speech recognition system would then seek to ensure a good match between person specification and technology specification, as well as a good fit between technology and application.

REFERENCES

[1] Simpson, C.A., McCauley, M.E., Roland, E.F., Ruth, J.C. and Williges, B.H., System design for speech recognition and generation, *Human Factors, 27,* 115-143, 1985

[2] Baber, C. and Noyes, J. Automatic Speech Recognition in Adverse Environments *Human Factors* (in press)

[3] Driskell, J.E. and Salas, E., Overcoming the effects of stress on military performance: human factors, training and selection strategies, In Ed. R. Gal and A.D. Mangelsdorf *Handbook of Military Psychology,* New York : Wiley, 1991

[4] Noyes, J. and Baber, C., Speech recognition in adverse environments: the role of human mediation, *Proceedings of the NATO / ESCA Workshop on Speech Under Stress*, Lisbon, 1995

[5] Baber, C., The effect of workload on the use of speech recognition systems, *Proceedings of the NATO / ESCA Workshop on Speech Under Stress*, Lisbon, 1995