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

The Effects of Workload on the Use of Speech Recognition Systems

Chris Baber

Industrial Ergonomics Group, School of Manufacturing & Mechanical Engineering,
University of Birmingham, Birmingham B15 2TT

ABSTRACT

The range of applications of speech technology has increased dramatically over the past two decades. With this increase, comes a new set of problems relating to the use of the technology. In this paper, one aspect of this set of problems is considered: workload.

The paper is divided into nine main sections. In the first two sections the aims of the paper and the topic of workload are introduced. In the third section, the effects of workload on human performance are reviewed. In the fourth section the status of workload as a stressor is considered. In the fifth section the effects of stress on performance are considered. In the sixth section, the production of speech is discussed and in the seventh section consideration is given to the possible effects of workload on the production of speech and on the use of speech recognition systems. In the eigth section the notion of workload is extended to consider the effect of physical workload on the use of speech recognition. Throughout the paper a number of problems will be highlighted and in the ninth section a number of proposals for solving these problems are presented.

1. INTRODUCTION

In recent years speech recognition technology has advanced to a stage where it is already being used in offices [1], in industry [2], and in military / avionic applications [3]. As the technology moves into the unforgiving 'real-world', the number and range of adverse factors increases [4]. It is proposed that a majority of problems in the use of speech recognition systems in the field arise from the interaction between person and recogniser. In some cases, this interaction can be defined with sufficient clarity to allow measures to be taken to deal with the problems. Thus, some of the adverse factors can be addressed from an engineering perspective (and are considered in other papers in these proceedings). In other cases, the problem itself is poorly defined and highly variable across individuals and contexts. One such problem is the subject of this paper and is known as workload.

Simpson et al. [5] suggest that the majority of successful applications of speech recognition have been in tasks which do not involve severe time pressure or threat to life. However, with the increase in interest in military applications for speech, this situation is changing. The US National Research Council [6] in a review of the effects of stress on the use of speech recognition notes that human speech is highly variable under stress and that many applications can be adversely affected by time demands, although the report does not consider workload in much detail.

Consider driving a car: in light traffic, one can drive the car and engage in conversation with passengers. As the volume of traffic increases or as the driving task becomes more difficult, the ability to hold a conversation while driving appears to decrease. In simple terms, this represents a focus of this paper: why do drivers stop talking when traffic becomes heavy, or, in broader terms, what effects do changes in workload have on people performing speech-based tasks ?

2. DEFINING WORKLOAD

In modern systems which are increasingly complex and dangerous, workload is becoming an important consideration. In many situations, there is the potential for demands to exceed operator capabilities and thus produce a decrement in performance (either in terms of reduced speed of response or in terms of human error). From the driving example given above, the increase in workload could result from an increase in the amount of information with which the driver has to deal (i.e., more cars), or from an increase in demands (i.e., more potential for collision), or from an increase in tasks (i.e., more gear changing), or from an increase in some other factor.

Norman and Bobrow [7] argue that as the 'difficulty' of a task increases, so too will the amount of effort which needs to be invested in performance. Thus, one would expect to see some indication of an increase in the level of 'effort' required to perform a task, perhaps through physiological changes or through a reduction in the ability to perform some additional tasks.

For the driving example, the need to maintain driving performance had an effect on the conversation, but could also have an impact on heart-rate or blood pressure.

In some tasks, performance will be resource-limited, i.e., influenced by the amount of resource (effort) a person can devote to task performance. In other tasks, performance will be data-limited, i.e., no matter how much resource a person devotes to a task, performance will remain at a constant level due to other limitations. For instance, in an air traffic control environment, an operator could, with effort, handle an increase in number of planes from one to three, but may not be able to cope with ten planes; indeed, performance may begin to deteriorate as more planes enter the frame.

3. WORKLOAD AND PERFORMANCE

There is a growing consensus of opinion in the human factors community that the level of workload will depend upon the relative cognitive demands made by competing tasks. The results from these studies are generally presented in the form of multiple resource theory (see figure one).

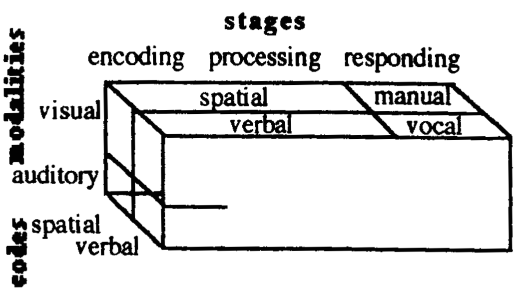

Figure one: Schematic diagram showing the proposed structure of multiple resources [after 8]

Broadly speaking, figure one should be read from left to right. In the first (encoding) stage, information from the world is presented in one of the codes and perceived using the appropriate sensory modality, e.g., a spoken message is transmitted using the verbal code and received by the auditory modality. The second (central processing) stage involves the interpretation of the information, again using the appropriate code (in this case, it will be the verbal code). The final stage involves defining an appropriate response to the information (in this instance, vocal).

This structure allows a number of hypotheses to be generated. First, if the code in which information is presented is appropriate to a modality, i.e., if it does not require translation, then response will be faster than if translation is required. Second, if a response is appropriate to the processing code, then response will be faster than if it is not. Both the first and second hypothesis can be seen from figure one by tracing the route for information within one of the lines, e.g., verbal information presented to the verbal modality processed verbally and requiring a vocal response. The third hypothesis is that, if more than one task presents information to one modality or requires processing with the same code or requires the same form of response, then there will a potential for workload to increase (with increases in information, task and demands). Given the notion of separate codes and resources, it is possible to make the further hypothesis that the information processing demands on a particular code would increase if more information was presented using that code, possibly leading to confusion.

Taylor [9] has noted that in high performance military aircraft, pilots' hands and feet are fully occupied with controls, vision is fully occupied with displays and with surveying the visual environment outside the aircraft and both reach and field of vision are severely limited. Consequently, efforts have been made to provide a means by which the pilot can communicate with on-board computers using the potentially spare speech modality. However, as Baber [10] has pointed out, speech is not always a 'spare' modality in the cockpit, with requirements for spoken communications between the pilot and a number of other people.

The main source of support for multiple resource theory comes from secondary task experiments. Participants are required to perform a primary task, such as controlling the movement of a cursor on the screen or 'flying' in a flight simulator, and then asked to perform a secondary task, such as entering specific information during the performance of the task. The manner in which the tasks are performed and information is presented to participants are varied across conditions in order to determine which combinations of spatial or auditory display and manual or spoken response leads to best performance. In general, the results tend to support the argument that performance on a verbal secondary task, when paired with a spatial primary task, is best when spoken responses are paired with auditory displays [11, 12, 13, 14]. Further, performance on a spatial secondary task, when paired with a spatial primary task, is best when manual responses are made to visual displays [13, 14]. Performance on a verbal secondary task is better than on a spatial secondary task, when paired with a spatial primary task [15, 16].

One can conclude that visual competition at input disrupts performance, especially if participants are required to perform visual scanning with tracking tasks. Spatial tasks show far more interference than combinations of verbal and spatial tasks. It is also possible that manual response in a secondary task can be paired with a manual primary task. This appears to contradict the premise of multiple resource theory. However, interference between tasks can be related to the timing of demands, as well as the code used. Thus, one could argue that competition for resources can be scheduled, depending on the relative urgency of the tasks.

Taken together the studies in multiple resource theory suggest that a viable model of human information processing can be developed which views information processing as related to the codes used in presenting information. Further, human information processing involves a limited attentional resource, employing discrete sensory modalities. There is some means of allocating 'resource' depending on task and demands, and a potential for overload when the demands exceed the capacity.

4. WORKLOAD AS STRESS

From physiological evidence, 'effort' can be seen as a generalised indication of arousal, i.e., increases in demand leads to increases in effort requiring increases in arousal [17]. One could assume that an individual will have a finite capacity for increasing arousal, and that beyond that capacity the individual will begin to suffer some form of stress or will seek to avoid further increase in effort, i.e., by stopping task activity.

The notion of a finite capacity for arousal is often presented as a limited pool of resources from which to allocate resource to different processing activities. As the pool decreases so the even distribution of the resource becomes more difficult. There seem to be three consequences of this state of affairs: cooperation, confusion and conflict.

In general, tasks which use similar information processing routines can cooperate in the performance of a task. For instance, a common finding is that simultaneous performance of two tasks involving tapping rhythms can be achieved relatively easily if the tasks share a common rhythm, but are more difficult to combine if they employ different rhythms. However, it is also evident that cooperation can be enhanced through training and practice. For instance, a skilled drummer can combine several rhythms played with hands and feet.

While pairing similar tasks can lead to cooperation, it can also lead to increasing demands on a particular resource. If one accepts the proposal that different types of information are processed using different processing codes, e.g., verbal and spatial, then it is possible to predict that two tasks requiring the same processing code will lead to either confusion or conflict. In the case of confusion, consider listening to two spoken messages; not only is this a difficult task, but also some of the information in one message might be interpreted as belonging to the other message. In the case of conflict, it will be impossible to perform the two tasks simultaneously and the person will need to give priority to one of the tasks.

5. STRESS AND PERFORMANCE

Baber and Noyes [4] suggest that there are broad categories of human performance which can be affected by stress and these are: psychomotor skills, memory functions, situational awareness, attention, planning, and judgement.

In terms of psychomotor skills, there are two major findings from the research: stress leads to an increase in the amount of time a person takes to respond to a signal, and extreme stress also results in an inability to act. In terms of memory functions, there is evidence to suggest that stress can disrupt recall from immediate memory. Stress also impairs the ability of people to extrapolate information from their surroundings, e.g. resulting in reduced efficiency in search behaviour. In terms of attention, stress restricts the number of cues to which a person can attend and reduces performance on vigilance tasks. Finally, in terms of general cognitive skills, stress results in a reduction in problem solving ability, performance rigidity (i.e., people stick to one strategy, even if it is not appropriate), and increased errors in following procedures. This latter finding is probably a combination of memory impairments (so that people do not remember where they are in a procedure) and performance rigidity (so that people ignore the procedure and adopt other strategies).

While the research tends to suggest a number of effects of stress, there is evidence that people can deal with stress. Wickens et al. [19] found that judgments requiring retrieval of information from long term memory were relatively unimpaired by stress. Thus, stress seems to result in impairments in performance for inexperienced people (for that task domain) drawing on relatively recently acquired information. A characteristic of behaviour under stress is a reversion to well learned behaviour patterns which are compatible with the situation in which one find oneself. This implies that some of the key problems resulting from stress can be reduced through the use of adequate training and through the design of systems which allow responses compatible to the situational demands.

6. THE PRODUCTION OF SPEECH

In order to determine the likely effects of workload on the use of speech recognition systems it is necessary to develop an idea of what activities are being performed when a person speaks to a device. A number of models of speech production have been proposed in recent years [20, 21, 22]. These models are sufficiently similar to allow them to be collapsed into a simple model (see figure two).

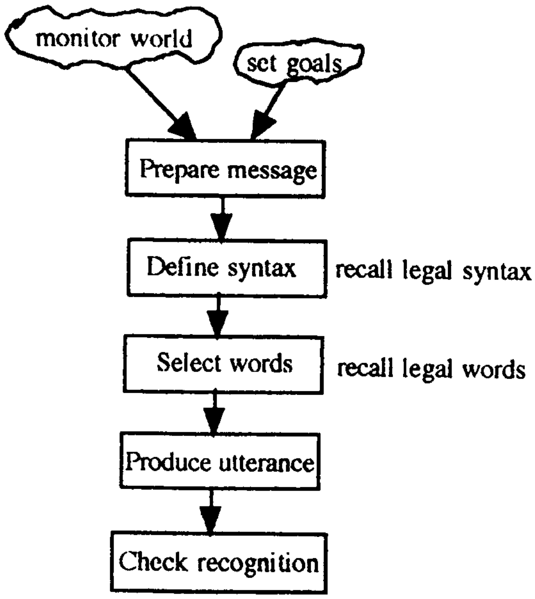

Figure two: Simple model of speech production

At the highest level, the meaning of the message to be communicated is generated. This level employs the speaker's knowledge of the language, the context of the conversation and the goals of the speaker to develop a basic intention to utter a message with a specific meaning. When using a speech recogniser, it is likely that the user will have defined goals, e.g., either to obtain information from a database or to issue a command to a system. At the next level, the syntax of the utterance is generated. With speech recognition, there is potential problem in the application of syntax, especially if the user is not aware at which level of syntax the recogniser is operating. Following the syntactic level, words are selected to fill the 'slots'. This is especially problematic for users of speech recognition as the vocabulary required may differ from the one that they wish to normally adopt. Finally, an articulatory program needs to be constructed and run to produce the utterance. At each level at set of 'rules' need to be invoked to govern the fluent performance of the activity at that level and the smooth transition to the next level. Further, the speaker will monitor what they say for any speech errors, and ascertain that the 'listener' has correctly heard the utterance.

7. WORKLOAD AND SPEECH

Having outlined a simple model of speech production, one can begin to map the possible sources of workload onto the stages of speech production. If one takes account of the peculiar nature of speaking to a computer, then it is possible to constrain the effects to only those which are conceivable in the use of speech recognition. In this section, the three major factors of information, task and demand load are considered. The amount of information, number of tasks, or demands will have a bearing on the ability of the individual to respond appropriately. The level of workload will interact with the person's ability to extract information from the environment; especially if the information requires translation, i.e., if it is presented to one modality but requires response using a different modality, e.g., a visual display requiring a spoken response.

7.1 Information load

One can map the possible effects of workload on performance outlined above onto these stages of speech production. In the first stage of speech production, in which an utterance is formulated, performance might be disrupted by several factors. At an extreme, the result could simply be no response at all. However, it is more likely that response will be delayed. The effect of this delay could be exacerbated by the time taken to process the speech by the recogniser. Thus, ***problem one***: time delays could play a critical role in the success of the speech application.

The second stage of speech production is the selection on an appropriate form for the words. In a number of applications, the syntax provided by the device is hidden from the user. This means that it is possible for users to mistake the syntax. This leads to ***problem two***: given the tendency to revert to well-learned behaviours, people may attempt to structure their utterances without due regard for the syntax.

The third stage of speech production involves the selection of words for the utterance. Graham and Baber [23] recorded a number of speech errors during a time-paced data entry task. The task involved entering information using the ICAO alphabet (A = alpha, B = bravo etc.). Many of the errors took of the form of substitution of other words for individual letters, e.g., 'Beta' for 'Bravo', or for letter sequences, e.g., 'John F Kennedy' for 'Juliet Foxtrot Kilo'. This leads to ***problem three***: speech errors, which involve violation of vocabulary constraints in using speech recognition systems, could arise because users revert to more familiar words and phrases. This problem will also be apparent to anyone who watches sports programmes: as the action becomes more exciting, commentators seem to revert to more and more cliches. One reason for this use of cliches could be to minimise the requirements to search for appropriate words while watching the changes in the action.

The fourth stage is the production of the utterance. At this stage, the various slips, slurps and stutters of speech errors will become apparent. While there is some indication that speech errors increase under stress, there has been no work on the relationship between speech errors and workload.

A further problem at the fourth stage occurs with monitoring the recognition of the speech recogniser. The user may misinterpret the feedback supplied by the speech recognition device. For instance, Usher and Baber [24] noted that, on some occasions, users appeared to regard textual feedback as a *general* indication that something had been recognised rather than as evidence of the recognition of a *specific* item. In other words, the users failed to monitor the meaning of the word and only responded to its presence. Given the probability that stress leads to a reduction in the number of cues people sample (and possibly the efficiency with which they sample cues), then it is probable that stress will exacerbate this effect. In effect, this can be expressed as ***problem four***: if the quantity of information in feedback increases, people will be more likely to treat this in general rather than specific terms. This will lead to an increase in the likelihood of user error (i.e., not responding to erroneous feedback).

7.2 Task load

A number of studies have shown that, when a secondary task is added to the use of speech recognition device, recognition accuracy tends to decrease [11]. What is interesting is that in several of these studies, the performance on the secondary task does not to be adversely effected by the use of speech. The suggests that, either the tasks can be paired optimally (as implied by multiple resource theory), or leads to ***problem five***: people vary their response strategy in terms of task load.

Further, one can also propose ***problem six***: if the number of tasks a person performs increases, there will be a tendency to narrow attention on specific cues, which may mean that feedback is treated in general rather than specific terms.

7.3 Demand load

There are several factors which come under the general heading of demand load; in this section attention shall be focussed on two: task complexity and time pressure.

Hapeshi and Jones [25] asked participants to perform a manual tracking task simultaneously with the use of speech recognition, and found a higher incidence of misrecognitions in the dual-task conditions. Furthermore, as the tracking task became more demanding, so the recognition accuracy decreased. While it is not possible to determine the precise cause of this decrease in recognition accuracy, one can propose two possible reasons. The first is a simple one of resource availability: as the tracking task took up more of the limited attentional resource, so there was less effort that could be directed at the maintenance of consistent speech. This explanation can be derived from multiple resource theory, especially when one notes that recognition feedback was presented visually. This leads to ***problem seven*** : performance will be impaired if feedback competes with the primary task modality. The second is one of time-sharing: as the tracking task took up more time, then speech production could have been 'rushed' top fit the available space. This latter explanation is supported by research into time-pressure and the use of speech recognition. For example, Simpson [26] found that recognition accuracy decreased on a time-pressured task because participants were shortening the pauses between words. When she played recordings of the speech to a connected word (as opposed to isolated word) recogniser, then recognition accuracy improved. Thus, ***problem eight*** : increased demands lead to an increase in speech rate.

A further problem with demand load is the disruption of short-term memory. This is demonstrated by the following study. Eight participants were trained on a paired-associate learning task. The participants were shown one word and then shown a second word to pair with the first. When prompted with one of the pair, they were then required to respond with the paired word. Participants responded in computer-paced and self-paced conditions (counter balanced across trials). The computer pacing presented words at a rate of one per second, while the self pacing allowed participants to work at their own rate. The task was performed on a Marconi MacroSpeak recogniser. Table one shows the percentages of correct responses for participants.

	practice	self	computer
correct	99	98	0.84
incorrect	0	2.5	6
no response	1.25	0	10

Table one: Percentage scores for each condition

There was a significant difference between self-paced and computer-paced performance for percentage of correct response and for percentage of no response. As the workload increased, participants either made mistakes or failed to respond. Thus, even under minimal levels of workload, people had problems in recalling a learned set of responses. Interestingly enough, there was no significant difference in recognition accuracy across the two conditions. If one recalls the example of speaking and driving given above, one can also propose that, ***problem nine***: as demand increases, people will reduce the amount of speech they utter, either by saying nothing or 'fitting' their speech to the demands of the situation.

8. EFFECTS OF PHYSICAL WORKLOAD

Some of the more successful applications of ASR in industry have been related to parts and baggage handling [Noyes et al., 2]. Given the physical nature of this work, one might expect there to be some effects of physical workload on the use of speech recognition. Any form of physical exertion will have a principal effect of altering breathing rate. Lierberman and Blumstein [27] note that people can compensate for such changes. However, it is likely that short term exertion involved in lifting packages will influence speech production. Thus, ***problem ten***: physical exertion involved in lifting and moving objects may influence the use of ASR.

An issue which can be considered related to physical workload is the possibility of fatigue arising from the prolonged use of ASR. For instance, Pallett [28] and Zarembo [29] note a decline a recognition accuracy over prolonged periods of ASR usage. Both authors suggest that this is due to vocal fatigue, with the voice tiring and so producing speech with changing characteristics. This argument is largely circumstantial and based on anecdotal evidence. It is also somewhat surprising when one considers that there are a number of jobs requiring people to talk for most of the working day, e.g., receptionists, telesales staff etc. The question then is whether people get tired when speaking to ASR devices.

Studies by Frankish et al. [30] suggest that prolonged use of ASR does indeed result in shifts in recognition accuracy. Their data indicate that the shifts pertain to overall recognition accuracy rather than variability. This implies that users of ASR go through a brief 'warm - up' period each time they use a device, with speech settling into an appropriate style after a period of a few minutes. The changes in recognition accuracy observed can be dealt with by simply retraining the device, either through adaptation or by the user re-enrolling vocabulary items. The subsequent drift from the re-enrolled templates was minimal.

9. DISCUSSION

Throughout the preceding sections, a number of possible problems have been identified in terms of the effects of workload on speech. Each problem can be tackled in a variety of ways and these will be considered in this discussion.

Problem one indicates that speech can be particularly problematic in time critical applications. While the processing time has been significantly reduced since the NRC [6] report, there is still good indication that the overall system response will be sluggish; primarily due to the fact that increased workload tends to delay response. Additionally, increased workload could lead to an increase in speech rate (problem eight). While this requires further examination, one could propose that speech be reserved primarily for data management and high-level control tasks. Finally, problem nine suggests that people may modify their speech rate to demands of the situation, e.g., when driving they may have long pauses between responses or between sentences. While this seems to contradict problem eight, it illustrates how speech is 'fitted' around the primary task and has implications for timing of spoken responses.

Problem two suggests that people might not be aware of the level of syntax that they are using. One could either seek to design the syntax to be easy to remember or seek to provide information to the user as to the syntax level, e.g., by showing the options available [Murray et al., 31].

Problem three suggests that people may introduce spurious words in the task. A solution to this problem is to consider the concept of habitability [Watt, 32], which argued for a fit between vocabulary and people's view of the task. It is interesting to note that there has been relatively little work on the concept of habitability since Watt's paper. Given the constraints imposed by speech recognition devices and the possible problems these might cause, it would seem that research on habitability is now highly desirable.

Problem four suggests that people can confuse or mistake the feedback that they are given. Studies suggest that one of the reasons for the failure to monitor feedback appropriately could be due to the lack of integration between task performance and feedback monitoring. Consequently, if feedback was incorporated into the primary task display [Baber et al., 33] or if feedback was presented in such a way as to interrupt primary task performance, e.g., through the use of synthesised speech [Frankish and Noyes, 34], then monitoring errors could be reduced to a negligible level. Of course, this brings us to problem seven, which suggests that feedback in one modality can interfere with the primary task (in this case, spoken feedback could disrupt the task of speaking a command). These findings suggest that, as with error correction, the provision of feedback has to be carefully considered and designed into the dialogue.

Problem five relates primarily to task load and requires investigation beyond speech technology, into the overall workload on the operator.

Problem six suggests that people may revert to a preferred strategy in high workload. There seem to be two solutions to this problem. The first is to see it as a training issue and to ensure that people are trained to only use one strategy. The second is a system design solution which is to ensure that only one strategy is encouraged and supported (given adequate interface design, it should also be possible to ensure that this strategy will be the most apparent to the user).

For problem ten, physical workload can be assumed to present a number of physiological effects. However, to date there is little research into their nature. While 'vocal fatigue' may be assumed to be a factor in the use of ASR, there appears to be little evidence to support its existence. Rather 'vocal drift' alters recognition accuracy; the 'drift' arises from users warming up and settling into an appropriate style of speaking. This means that problems can be dealt with relatively easily by retraining vocabulary items.

There is some agreement that compatibility can reduce the detrimental effects of workload. For instance, in tasks involving a high level of visual activity, it might be possible to reduce demands by using the 'free' speech channel to 'off-load' some of the demands. In practice, this often leads to calls for the introduction speech technology into the operating environment, e.g., the cockpit. However, without full exploration of the tasks demands in the environment and without full consideration of alternative solutions to the problem of workload, one might make matters worse rather than better. For example, there might already be a high level of verbal communication in the environment, and introducing speech to reduce spatial demands may only shift the locus of demands to verbal tasks.

10. REFERENCES

[1] Noyes, J.M. and Frankish, C.R., A review of speech recognition applications in the office, *Behaviour and Information Technology, 8,* 475-486, 1989

[2] Noyes, J.M., Baber, C. and Frankish, C.R., Industrial applications of ASR *Journal of the American Voice I/O Society, 12,* 51-68, 1992

[3] Moore, T.J., Speech technology in the cockpit In Ed. R.S. Jensen *Aviation Psychology* Aldershot: Gower Technical Press, 1989

[4] Baber, C. and Noyes, J.M., Automatic speech recognition in adverse environments, *Human Factors,* in press

[5] 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 (2)* 115-141, 1985

[6] National Research Council, Committee on Computerized Speech Recognition Technologies, *Automatic Speech Recognition in Severe Environments* Washington, DC: National Research Council, Commission on Engineering and Technical Systems, 1984

[7] Norman, D. and Bobrow, D., On data-limited and resource-limited processing, *Journal of Cognitive Psychology, 7* 44-60, 1975

[8] Wickens, C.D., *Engineering Psychology and Human Performance,* New York: Harper Collins, 1992

[9] Taylor, R.M., Integrating voice, visual and manual transactions: some practical issues from aircrew station design In Ed. M.M. Taylor, F. Neél and D.G. Bouwhuis *The Structure of Multimodal Dialogue* Amsterdam: North-Holland, 1989

[10] Baber, C., *Speech Technology in Control Room Systems: a human factors perspective,* Chichester: Ellis Horwood, 1991

[11] Harris, S.D., North, R.A. and Owens, J.M., A system for the assessment of human performance in concurrent verbal and manual control tasks, *Behaviour Research Methods and Instrumentation, 10,* 329-333, 1978

[12] Wickens, C.D., The structure of attentional resources, In Ed. R. Nickerson *Attention and Performance VIII,* Hillsdale, NJ: LEA 1980

[13] Wickens, C.D. Sandry, D.L. and Vidulich, M., Compatibility and resource competition between modalities of input, central processing and output, *Human Factors, 25,* 227-240, 1983

[14] Wickens, C.D., Vidulich, M. and Sandry-Garza, D., Principles of S-C-R compatibility with spatial and verbal tasks: the role of display location and voice interactive display control interfacing, *Human Factors, 26,* 533-542, 1984

[15] Wickens, C.D. and Liu, Y., Codes and modalities for multiple resources: a success and some qualifications, *Human Factors, 30,* 599-616, 1988

[16] Baber, C., Stammers, R.B. and Taylor, R.G., An experimental assessment of automatic speech recognition in high cognitive workload situations in control rooms, *Proceedings of the 11th. International Ergonomics Association Conferences,* London: Taylor and Francis, 833-835, 1992

[17] Sanders, A.F., Towards a model of stress and human performance, *Acta Psychologica, 53,* 61-97, 1983

[18] Wickens, C.D., Stokes, A.F., Barnett, B. and Hyman, F., The effects of stress on pilot judgement in a MIDIS simulator, In Ed. O.Svenson and J.Maule, *Time Pressure and Stress in Human Judgement and Decision Making*, Cambridge: Cambridge University Press, 1991

[19] Garrett, M.F., Production of speech: observations from normal and pathological language use, In Ed. A.W.Ellis, *Normality and Pathology in Cognitive Functions*, London: Academic Press, 1984

[20] Mackay, D.G., The problems of flexibility,m fluency and speed-accuracy trade-off in skilled behaviour, *Psychological Review, 89*, 483-506, 1982

[21] Dell, G.S., A spreading activation theory of retrieval in sentence production, *Psychological Review, 93*, 283-321, 1986

[22] Graham, R. and BABER, C., User stress in ASR In Ed. E.J. Lovesey London: Taylor and Francis, 463 - 469, 1993

[23] Usher, D.M. and Baber, C., *Automatic Speech Recognition in Grid Control Rooms*, Bristol: CEGB Report TD/STM/89/10081/N, 1989

[24] Hapeshi, K. and Jones, D.M., Concurrent manual tracking and speaking: implications for automatic speech recognition, In Ed. M.J. Smith and G. Salvendy, *Work with Computers: Organizational, Stress and Health Aspects*, Amsterdam: North-Holland, 412-418, 1989

[25] Simpson, C.A., Speech variability effects on recognition accuracy associated with concurrent task performance by pilots, *Ergonomics, 29*, 1343-1357, 1986

[26] Zarembo, L.A., The usefulness of voice recognition on the floor of the New York Stock Exchange, *Proceedings of Speech Tech'86*, New York: Media Dimensions, 69-72

[27] Frankish, C.R., Jones, D.M. and Hapeshi, K.C., Maintaining recognition accuracy during data entry tasks using speech input, In Ed. E.J. Lovesey, *Contemporary Ergonomics 1990* , London: Taylor and Francis, 445-453, 1990

[28] Watt, B., Habitability, *American Documentation*, 1968

[29] Baber, C., Usher, D.M., Stammers, R.B. and Taylor, R.G., Feedback requirements for automatic speech recognition in the process control room, *International Journal of Man-Machine Studies, 37*, 703-719, 1992

[30] Frankish, C.R. and Noyes, J.M., Sources of human error in data entry tasks using speech input, *Human Factors, 32*, 697-716, 1990