Proceedings of the Institute of Acoustics

THE DEVELOPMENT OF A COMPUTER BASED CLINICAL SPEECH AUDIOMETRIC TEST

C J James and J M Bowsher

University of Surrey, Department of Physics, Guildford, Surrey.

1 INTRODUCTION

Previous investigation into the properties of "phonetically balanced" wordlists (PB lists) has yielded useful insight into the intelligibility characteristics of individually presented mono-syllables (James et al., 1990). Words whose subjective intelligibility was impaired by certain presentation conditions were termed "fragile", whereas those whose intelligibility was not functionally affected were termed "robust". These properties can be explained in terms of the relative familiarity of the objects coupled with their physical and phonetic structure. It was conjectured that the notions of robustness and fragility could be used in the design of a speech audiometric test, providing useful information in clinical diagnosis.

Initial work went towards the design of a prototype test which could be used to obtain a fast measure of a subjects hearing sensitivity with respect to speech. This would provide information in a similar way to the conventional measurement of speech reception threshold (SRT), which is defined by presentation level at which an individual obtains a 50% score with PB lists.

Experimentation was required to gain information on the intelligibility characteristics of a sample of mono-syllables taken from existing clinical speech material. The experimentation incorporated a novel adaptive paradigm as used in the final prototype test. This paradigm used the results from pairs of presented words to make decisions about the adaptation to be applied to the presentation conditions. In the prototype test the paradigm was more important in that each pair of words presented consisted of one "level fragile" word and one "level robust" word.

The paradigm, and a 6 dB decrement/increment adaptation step, were chosen on the basis of analysis of iterative computational simulation of subjective experimentation using individual word intelligibility models as described below.

The results from the subjective experimentation allowed choice of sets of "robust" and "fragile" words for use in the prototype test. Several trials of the prototype test were performed using subjects of various ages.

2 COMPUTER MODELLING

2.1 Principles

The research and experimentation towards the creation of a self-adaptive test itself incorporated a self-adaptive approach. The first stage of the experimentation involved speculation on how the final test would utilise sets of words with particular properties. From the study of non-equivalence of isophonemic word lists (James et al., 1990) it was discovered that certain words remained reliably intelligible under a particular range of conditions, whereas there were also words whose intelligibility was generally inconsistent, and perhaps there were some words whose intelligibility was adversely affected by the various conditions. It was proposed that for example, certain words were "robust" to certain conditions whilst some would be "fragile", there would also be a set whose intelligibility was inconsistently affected by the conditions, these were labelled "unreliable". It was decided to tackle robustness and fragility in terms of the effects of presentation level on the intelligibility of real speech. The presentation level parameter is most frequently used in audiological assessment in order to determine the speech reception threshold (SRT) of the subject.

The presentation level where a subject correctly responded to a "level robust" word and incorrectly to a "level fragile" word should represent the speech reception threshold, provided that the properties of robust and fragile words are well defined (see figure 1). If the test recipient suffered from a certain non-linear hearing loss, such as high frequency loss or recruitment, the not only would the obtained threshold be elevated, but the functioning of the test could be affected. However, if the test could be said to function "normally" with normal hearing subjects and "abnormally" with subjects suffering from deficit the test would be diagnostically useful in that it could be used distinguish between the two groups. The principle of the test design is to "home in" on the recipient's threshold with respect to level. This is achieved by adaptation of the presentation level until the threshold (SRT) condition is satisfied.

2.2 Use of intelligibility models in iterative computational experiments

The first stage of experiment involved modelling in some way the information that it was hoped was to be gained about the test set of real words. The intelligibility of the one hundred and twenty words to be used in the subjective experiments was modelled by the generation of several sets of mean thresholds each with associated variance in these thresholds. The intelligibility characteristics of each word were thus represented by two numbers; The a threshold mean value, and a value representing the variance or width of that value. Actual sets of data were generated randomly with rectangular distribution. Each data set had a defined range of mean threshold accompanied by a range of variance of thresholds. There were three ranges used for each parameter to make nine intelligibility models in total:

		Threshold Ranges/dB		
		a. 20-29	b. 15-34	c. 10-39
Variance Ranges /dB	x. 0-6	ax	bx	cx
	y. 0-12	ay	by	cy
	z. 0-18	az	bz	cz

Table I. Combination of two sets of parameters to make nine models of word intelligibility in total.

2.3 Adaptive Paradigms

Three adaptive protocols were considered. Two of these used a two-by-two rule such that the results obtained from a pair of presented words were used in choosing the modification of presentation conditions and were in keeping with the proposed adaptive rule that would be used in a final test. The adaptation rules are as follows,

Both words correct/intelligible - decrement level by one step

One correct, one incorrect - do not change level

Both incorrect/unintelligible - increment level by one step

the only difference between these two protocols being the size of the step, 3dB or 6dB. The other adaptive protocol was consistent with the "Manchester Method" where a correct response results in the parameter being changed by two steps negatively (-6dB) whereas an incorrect response results in a single step positive increase (+3dB) in the parameter. In order to assess which protocol would produce the "best" results in the final real experiment all the intelligibility models described above were used.

2.4 Iterative Computational Trials

A real trial was simulated by comparing the current presentation level with the model value in combination with plus or minus a randomly generated quantity up to the value of the associated model variance. If the presentation level was higher or equal to the model threshold value then the trial response was positive, if lower then negative. The results from a trial were then used to adapt the presentation level according to the protocol being used. The order of "presentation" was randomised in keeping with the subjective study so as to provide as faithful a simulation as possible. Other parameters such as the starting level in relation to the mean thresholds of the models were chosen as much as possible to be in keeping with further subjective trials.

A "model" experiment was run using forty iterations for each protocol in combination with each model intelligibility set. This simulated the use of forty subjects in a real study. The data set for each model consisted of a set of trial outcomes linked to presentation level for each of the 120 words. With forty iterative experiments this produced forty data points for each word.

Using auxiliary programs the intelligibility function of each word could be assessed in several ways. A mean threshold of intelligibility could be obtained either from the simple mean (average) of the highest level at which the word was incorrectly perceived (maxlw) and the lowest level the word was correctly perceived (minlr), or by linear regression of all the data points to obtain a level at which the word crossed the boundary from intelligibility to non-intelligibility (intercept). Similarly a measure of the variance of the intelligibility of each word could be obtained from looking at the difference (maxlw-minlr), called difference, and the slope of the line obtained by linear regression (gradient).

3 SUBJECTIVE EXPERIMENTATION

3.1 Use of results from the model experiments in the "real" experiment

By looking at the efficiency of any particular paradigm in providing results that provide a measure of the original intelligibility model parameters for each word, it is possible to choose the most "efficient" paradigm. The 6 dB decrement/increment two-by-two paradigm was used in the real subjective experiment in that the computational trials showed it to provide the best results with the least number of iterations, over all the intelligibility model sets.

The order of presentation of words from the set of one hundred and twenty was randomised for each subject. The results using any number of subjects could be analysed by an auxiliary program to produce a running impression of results. Thus the results from subsequent experiments could be compared giving an indication of whether any great gain could be made in using a further number of subjects. Eventually seventeen subjects took part in the experimentation.

Figure 2 shows the auto-correlation for the intercept parameters for one of the intelligibility models (two iterative experiments) and the experimental data, bottom and top respectively, versus iteration or experiment. The "ay" or 10-29 dB thresholds, 0-12 dB is chosen for comparison on the basis that the model and the iterative results obtained from the model were fairly similar to the experimental set of results. The overall trend for both this particular model and the experimental data is of convergence toward some final exact values for the parameters.

4 THE TEST PROTOTYPE

4.1 The use of experimental data in the test prototype
The principle aim of the experimentation was to gain information about the intelligibility characteristics of each word in the test set. This information, in a raw form, was a set of data points, each point consisting of a presentation level paired with the outcome of a trial. These data could also be reduced to four parameters: The two parameters intercept and average representing directly the mean intelligibility threshold of a word, the parameters gradient and difference are indicative of the consistency of this mean intelligibility threshold.

For the purposes of the design of the test it was required that words be chosen from the full set such that they could be considered reliably "robust" or reliably "fragile" at threshold. To determine a working threshold a mean was taken from the mean of presentation levels used in the experiments. A requirement of the obtained fragile words was that they each became to a good degree intelligible a few steps above threshold and similarly the robust words should become unintelligible below threshold (figure 1). Thus the adaptive process should function properly in the running of a final test. Words were categorised according to their intelligibility properties around the threshold level. Various values of parameters could be applied to select sets of words that would be suitable for a prototype test. Changing the values of the selection parameters could produce a range of groups of words which could either be too sensitive, too insensitive or suitable for the correct functioning of the test.

4.2 Informal trials
Informal trials have been run using a test prototype with words obtained using particular selection parameters. Several "normal" subjects have been used both with and without knowledge of the original experiment along with several older subjects. So far with the "normal" subjects the test ran very well with a stable threshold value being obtained (figures 3, 1). The test proved less stable with older subjects (figure 3, 3), but still managed to "hover" around a threshold value. We may make the conjecture that mild high frequency loss, likely in the older subjects, could cause this mild instability, and that this may be used in the detection of high frequency loss. It would be possible to look at these problems by experimentation with normal subjects using filtered speech material coupled with observations made with a group of older and/or hearing impaired subjects.

4.3 Further development
The design of a similar kind of test has been proposed to look at the "roll-over" effect which is seen in the results from speech tests obtained with subjects with particular kinds of dysfunction. A record is made of the maximum fractional score attainable by the subject at an optimum level. Any deviation in presentation level from this optimum value results in a loss in speech recognition score. Lowering the level reduces the "strength"

of stimuli presented to the subject whilst increasing the level "overloads" the subjects auditory system. This effect is quite often seen in patients with recruitment problems caused by pathologies like acoustic neuroma (Jerger & Jerger, 1984), and provides a powerful indication to the presence of such.

It is proposed to use a self-adaptive protocol to probe the upper threshold of speech reception using carefully chosen stimuli as in the downward search test. Use may be made of results obtained from clinical investigation of patients with these kind of problems. It is conceivable that the physical structure of particular words will render them sensitive to this kind of hearing problem: The presence of a strong vowel sound overbearing the weaker consonantal sounds in combination with a relatively unfamiliar stimulus may make recognition more difficult for the recruiting subject.

5 CONCLUSIONS

This work has so far only scratched at the surface in incorporating the notion of robust and fragile objects in audiometric testing. We have so far achieved in this prototype a method of obtaining a value of speech reception threshold within presentation of about ten to sixteen words, using conventional methods this would require presentation of at least thirty words. The test was administered using a computer, also allowing easy display of results.

It is hoped that in future work we may address the acquisition of further knowledge required in the diagnosis of auditory dysfunction, using similar adaptive techniques, reducing the time required for speech audiometric testing in the clinical environment.

6 REFERENCES

James, C.J., Bowsher, J.M. and Simpson, P.J. (1990) "Digitisation effects and the non-equivalence of isophonemic word lists: A cautionary tale." Submitted for publication to Acustica December 1989.

Jerger, S. and Jerger, J. (1981) "Acoustic Schwannoma" in Auditory Disorders, Chpt 1. Little, Brown and Co., Boston.

SPEECH AUDIOMETRY

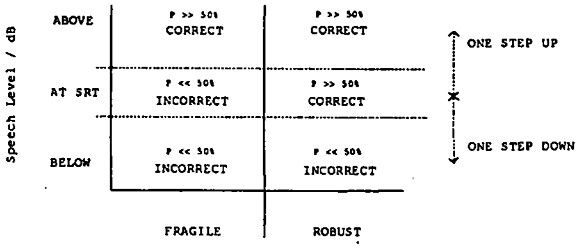

FIGURE 1. DERIVATION OF SPEECH RECEPTION THRESHOLD (SRT), USING ROBUST AND FRAGILE WORDS

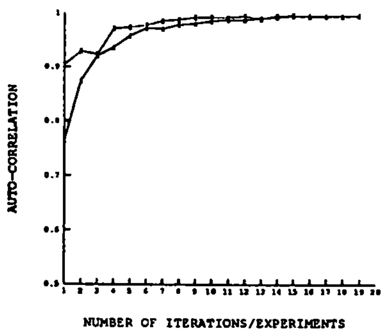

FIGURE 2. AUTO-CORRELATION OF RESULTS WITH ITERATION/EXPERIMENT

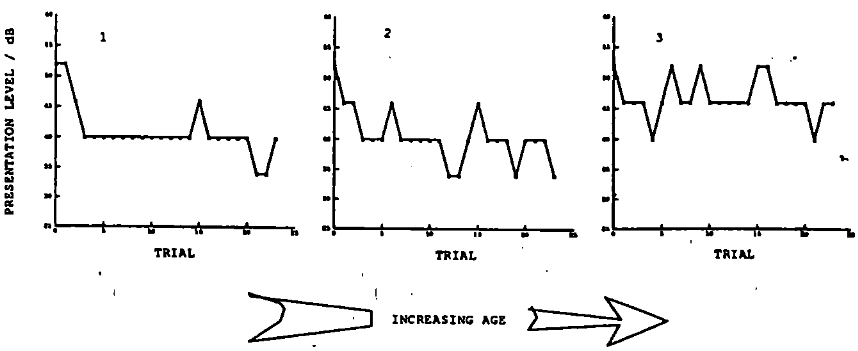

FIGURE 3. PRESENTATION LEVEL VERSUS TRIAL USING TEST PROTOTYPE.